New Methods for Studying

Structure and Dynamics of RNA

Een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

Ter verkrijging van de graad doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 7 december 2011
om 10.30 uur precies
door
Ramon Machiel van der Werf
Geboren op 26 januari 1978
te Leeuwarden

Promotor:	Prof. dr. S.S. Wijmenga
Copromotor:	Dr. M. Tessari
Manuscriptcommissie:	
	Prof. dr. G. Pruijn
	Prof. dr. L.M.C. Buydens
	Dr. R.M. Scheek

R.M. van der Werf, New Methods for Studying Structure and Dynamics of RNA
Ph.D. Thesis, Radboud University Nijmegen, The Netherlands
With summary in dutch
ISBN/EAN: 978-94-6191-064-6
Printed by Ipskamp Drukkers, Enschede

Table of Contents

Used abbreviations

AL – Apical Loop

ALD – Apical Loop of the hepatitis B virus of Duck

ALH - Apical Loop of the hepatitis B virus of Human

BMRB – Biological Magnetic Resonance data Bank

cccDNA – Covalently Closed Circular Deoxyribose Nucleic Acid

CHEOPS – CHEmical shift de novO structure derivation Protocol employing Singular value decomposition

CPMG – Carr-Purcell-Meiboom-Gill (train of 180° pulses)

DHBV - Hepatitis B virus of Duck

DNA – Deoxyribose Nucleic Acid

DQF-COSY – Double Quantum Filtered COrrelation SpectroscopY

DR1 – Direct Repeat 1

EDTA – EthyleneDiamineTetraAcetic acid

EM – Electron Microscopy

ESU – ElectroStatic Unit

EV – Entero-Virus

FED – Full Epsilon of the hepatitis B virus of Duck

FEH - Full Epsilon of the hepatitis B virus of Human

GAF – Gaussian Axial Fluctuation

HBV – Hepatitis B Virus

HCV – Hepatitis C Virus

HIV – Human Immunodeficiency Virus

Hsc – Heatshock constitutive protein

Hsp – Heatshock protein

HSQC – Heteronuclear Single Quantum Coherence

IRES – Internal Ribosome Entry Site

LAB – LABoratory axis frame

LAF – Local Axis Frame

MAF – Molecular Axis Frame

mRDC – magnetic field alignment induced Residual Dipolar Coupling

mRNA – messenger Ribo Nucleic Acid

NMR – Nuclear Magnetic Resonance

NOE – Nuclear Overhauser Effect

nt - nucleotides

nRTI – nucleoside analog Reverse Transcriptase Inhibitor

ORF – Open Reading Frame

PAF – Principal Axis Frame

PB – Primer bulge

PBD – Primer Bulge of the hepatitis B virus of Duck

PBH - Primer Bulge of the hepatitis B virus of Human

PCS – PseudoContact Shift

PDB – Protein DataBank

pgRNA – pregenomic RNA

PRE – Paramagnetic Relaxation Enhancement

RCSA – Residual Chemical Shift Anisotropy

RDC – Residual Dipolar Coupling

RH – RNAse-H

rMD – restrained Molecular Dynamics

RMSD – Root Mean Square Deviation

RNA – Ribose Nucleic Acid

RNAi – Ribose Nucleic Acid interference

RT – Reverse Transcriptase

SELEX – Systematic Evolution by Ligands by EXponential enrichment

SVD – Singular Value Decomposition

TAD – Torsion Angle Dynamics

TAR - Trans-Activation Response element

TP – Transcription protein domain

tRNA – transfer RiboNucleic Acid

Tyr - Tyrosine

WHO – World Health Organization

Notations

$\mathbf{A}$ – Matrix or Tensor

A_{ij} – Matrix or Tensor element, column i, row j

β – Variable

$\hat{H}$ – Operator

$\hat{\hat{R}}$ – Superoperator

(#) – Reference

(#.#) – Equation number

Chapter 1

General Introduction

1 General Introduction

1.1 Outline of thesis

The research described in this thesis falls into the field of NMR structural biology of nucleic acids and covers two main subjects. a) Development of new NMR methods for the determination of structure and dynamics of nucleic acids. b) Application of these and other established methods to two related bio-systems of high biomedical interest, namely the 62 nucleotide ε RNA elements of human and duck hepatitis B virus. These RNA elements are essential for the replication of the hepatitis B virus.

Chapter 1 provides background and introduction into NMR structural biology of RNA. In this chapter, after a brief prelude on the main biological roles of RNAs, the epidemiological, molecular biology, and structural aspects of the hepatitis B virus are reviewed. Second, the structure of RNA (and DNA) are described. Subsequently, NMR structure determination of nucleic acids is reviewed, paying specific attention to residual dipolar couplings and isotropic proton chemical shifts. A formal theoretical derivation is given for the dependence of residual dipolar couplings and residual chemical shift anisotropy on structure parameters of partially aligned biomolecules. In addition, the theoretical background to isotropic chemical shifts is described. Finally, internal dynamics in nucleic acids are discussed and how they can be characterized using NMR spin relaxation and/or residual dipolar couplings.

In **chapter 2**, the structure and dynamics of the apical stem with its loop of the ε element of the hepatitis B virus of the duck are described. These have been solved by ‘classical’ NMR methods. The methods yield more insight into how this element performs its function in the replication cycle of the hepatitis B virus of the duck.

In **chapter 3**, the assignment of NMR spectra of the primer-binding loop of the ε encapsidation signal of the duck hepatitis B virus is presented. This was a first step in obtaining NMR parameters required to solve the structure of the primer-binding loop of the duck.

In **chapter 4**, the solution structure and dynamics of the full 62 nt ε encapsidation elements of the duck and human hepatitis B virus are presented, discussed, and compared. Many characteristics of the HBV replication mechanism had previously been worked out based on a fully functional in-vitro replication system available for the duck hepatitis B virus and a partially functional in-vitro replication system recently set up for the human hepatitis B virus. To understand the molecular details of the replication function of ε, it is also crucial to have structure and dynamics information available. The presented study provides the first description of a complete structure and dynamics of such large RNAs.

In **chapter 5**, a new NMR structure derivation methodology for nucleic acid helices that employs non-exchangeable proton chemical shifts as a source of experimental structure restraints, is introduced and demonstrated on two A-helix RNA structures. The method makes use of the helix parameter description of the three-dimensional structure of nucleic acids. The proton chemical shifts are found to define with high accuracy the translational and rotational basepair and basepair step parameters. To the best of our knowledge, this is the first time that chemical shifts are used as the sole source of structural information for deriving nucleic acid helix structure.

In **chapter 6**, the method presented in chapter 5 is used to determine the Dickerson dodecamer B-DNA structure from its published experimental proton chemical shifts. The resulting structures are compared to X-ray structures. We find that the resulting helix structures are defined with high precision and closely resemble the X-ray structures. The structures based on proton chemical shifts show no NOE, dihedral angle and/or RDC violations. In summary, the Dickerson dodecamer structures based on proton chemical shifts is as precise as the NMR structure with a large number of restraints.

Finally, in **chapter 7,** the main results presented in the thesis are summarized.

1.2 RNA and its role in biology

The biological relevance of RNA molecules is illustrated by the history of Nobel prices awarded for research on RNA (DNA) over the past fifty years (11). A summary of these awards can be found in Table 1.1, of which some will be highlighted here.

Many prices were awarded for the discovery of the functioning of RNA as a messenger molecule in the expression of the genetic code, i.e. protein synthesis. In 1989, S. Altman and T.R. Cech received the Nobel price for chemistry for “their discovery of catalytic properties of RNA”. S. Altman found that RNAse P, an enzyme consisting of both amino acids and RNA, loses its catalytic activity upon removal of the RNA. Cech discovered that one RNA strand could cleave another. Thus, RNA is not only a mere messenger, but can also act as an enzyme, a ribozymes. Previously to this work, proteins were known as the only biomolecules showing enzymatic activity.

Table 1.1. Summary of the Nobel prize winners associated with RNA (years 1959-2009)

Year	*Who*	*Nobel price for*
1959	S. Ochoa, A. Kornberg	discovery of the mechanisms in the biological synthesis of ribonucleic acid and deoxyribonucleic acid
1962	J.D. Watson, F.H.C. Crick, M.H.F. Wilkins	discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material
1968	M. Nirenberg, R.W. Holley, H.G. Khorana	interpretation of the genetic code and its function in protein synthesis
1989	S. Altman, T.R. Cech	discovery of catalytic properties of RNA
1993	R.J. Roberts, P.A. Sharp	discoveries of split genes, i.e. gene splicing
2006	A.Z. Fire, C.C. Mello	discovery of RNA interference, gene silencing by double-stranded RNA
2009	V. Ramakrishnan, T.A. Steitz, A.E. Yonath	studies of the structure and function of the ribosome

Another important discovery by A.Z. Fire and C.C. Mello, who received the Nobel price in chemistry in 2006, was the discovery of RNA interference. RNA interference is a gene suppressing mechanism in which mRNA undergoes base pairing with a complementary short RNA strand (RNAi); this RNA complex is cleaved with the help of associated proteins, and in that way interferes with the expression of mRNA. Cell proliferation, development, immunity, and tumor genes can be influenced by means of RNA interference (12).

In 2009, Ramakrishnan, Steitz and Yonath received the Nobel price for their work on the three-dimensional structure of the ribosome, a large complex structure consisting of proteins and RNAs. The crucial role of RNA is illustrated by the structure of the ribosome, which shows that its catalytic activity needed for translation fully derives from its RNA constituent (13; 14).

Today, the human genome as well as that of a growing number of other species has been charted. Remarkably, the human genome contains only a relatively modest number of genes (ca 30 000) and is now known to consist in large part (circa 98 %) out of well-conserved "non-(gene-)coding" DNA that often codes for regulatory RNAs (15). Indeed, in the modern field of RNA research many new RNAs are found, often with a regulatory function. For instance, the discovery of riboswitches (16; 17), which act as regulatory RNA elements for transcription or translation. As the concentration of a ligand in a cell increases, for example guanine, the ligand binds to the riboswitch. The riboswitch then undergoes a conformational change, which in turn results in suppression of transcription or translation of a particular gene. The Nobel prices, including the discovery of catalytic RNAs, together with the recent discoveries of regulatory RNAs, such as riboswitches and iRNA, illustrate the remark S. Ochoa already made in his Nobel lecture (1959): "these particles [RNAs] are at the threshold of life and appear to hold the clue to a better understanding of some of its most fundamental principles". This is also illustrated by the fact that RNA is often an important element in viral replication (18), like for instance in that of the hepatitis B virus.

1.3 The Hepatitis B virus

The hepatitis B virus (HBV) is a member of the family of hepadnaviridae (6; 19). a family of small enveloped DNA viruses infecting mammalian and avian hosts. Members of this family replicate via an RNA intermediate. HBV has a very small genome consisting of only ~ 3 kb double-stranded DNA. Consequently, the genome contains many overlapping open reading frames (ORF) as depicted in Figure 1.1. The ORFs of S, pre-S1, pre-S2, C, pre-C, and X overlap with the ORF of the multi-domain protein P. One of the domains of P is the viral reverse transcriptase that transcribes the (pregenomic) mRNA back into genomic DNA. Therefore, protein P plays a key-role in the viral replication, as will be discussed later.

HBV is expected to be less prone to mutations than HIV and tuberculosis (20; 21). This is partly due to overlapping ORFs in the genome of the hepatitis B virus and contains a minimal amount of non-coding DNA (22). In addition, inside the capsid there is no space for much more than the DNA of the virus plus its essential proteins (19).

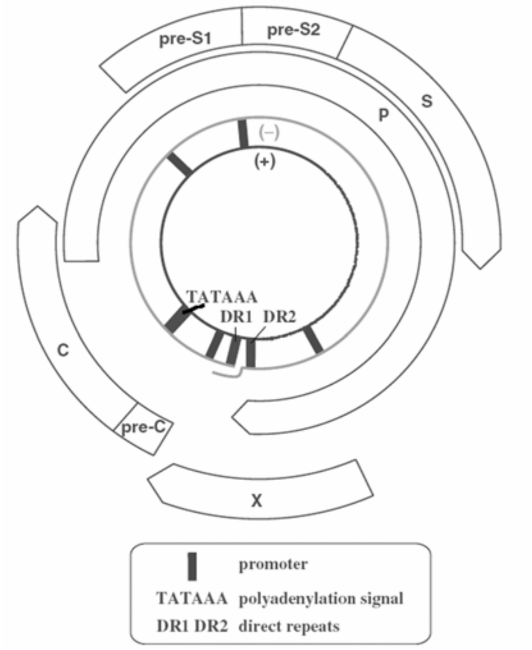


Figure 1.1: Hepatitis B virus plasmid. Overlapping Open reading frames are indicated on the double stranded DNA (+ and – strand). The letters represent the following ORFs: C and pre-C represent the capsid protein ORFS, S, pre-S1, and pre-S2 for viral envelope protein ORFs, and X is a transcription factor ORF, P is an ORF that encodes for the polymerase. (6).

1.3.1 Epidemiology

Hepatitis B can be transmitted between humans by contact with blood or other body fluids, i.e. semen and vaginal fluid (23; 24). These modes of transmission are thus similar to that of the human immunodeficiency virus (HIV) (23). Unlike HIV, HBV can survive outside the body for at least 7 days and can still cause infection (23; 25). Symptoms of a HBV infection are yellowing of skin and eyes (jaundice), dark urine, extreme fatigue, nausea, vomiting and abdominal pain. In spite of the absence of a medicine against HBV infection, 95 % of the persons recover from it without any medication and 5 % of the infected persons develop a chronic infection (2). Chronic infection can lead to liver damage (2) . For instance, about 25 % of the adults who become chronically infected during childhood later die from liver cancer or cirrhosis (scarring of the liver) caused by the chronic infection (23).

The World Health Organization fact sheet 204 (23; 25), reports that about 2 billion people are infected by HBV and more than 350 million of these have chronic infections. Because of the number of infected people and the absence of an effective treatment, chronic infection forms a major global health problem. The WHO fact sheet shows a high level of occurrence of (chronic and acute) infection in the developing world, sub-Saharan Africa, most of Asia and the Pacific. Liver cancer caused by HBV is the major cause of cancer in women and is among the first three causes of death in men. The impact of HBV infections is further illustrated by a comparison with HIV infection. The WHO estimates, that at the end of 2007, 33.2 million HIV infections occur worldwide, which is only one-tenth of the number of hepatitis B infections. For tuberculosis, a similar number of infections as for HBV are reported in 2007. The WHO claims that HBV is 50 to 100 times more infectious than HIV (23).

Vaccination against HBV infection is possible as a preventive measure; the vaccine is a solution of coat proteins from the viral envelope (antigens) of the virus (6; 26-28). One of the reasons why a vaccine could be developed against HBV but not yet against HIV, is that HBV replicated slower and thus is less prone to mutations than HIV (29). The vaccine against HBV has an efficacy of 95% and vaccination of newborns from infected mothers has an efficacy of at least 75 % (24; 30). 5% of the population does not respond to vaccination (24). Vaccination does not provide long-term protection and requires repeated vaccination every ~25 years (30). To prevent HBV from occurring, a global vaccination is needed. This requires special care for those not sensitive to regular HBV vaccines. To reach these ‘non-responders’ new vaccines need to be developed. Changes in vaccine production systems and in formulations would take the vaccine beyond economic affordability (24).

For many years, HBV treatment was mainly the administration of interferon alfa. Interferon alfa is a protein that induces an immune response. Side effects of this therapy (fever, myalgias, thrombocytopenia, and depression) make it a difficult treatment for many patients. Moreover, in many patients a flare of liver injury occurs during administration of interferon alfa. Although sometimes disquieting to patients and physicians alike, these flares are intrinsic to the therapy and, as markers of enhanced antiviral immune responsiveness, often presage a successful outcome. However, treatment with interferon Alfa is generally contraindicated in very advanced liver disease, since in such cases the flares may precipitate overt liver failure (2). These contraindications make it difficult to treat patients with hepatitis B.

In 2002, a true anti-viral alternative to interferon Alfa treatment became available, namely Lamivudine (31). Lamivudine belongs to the class of nucleoside analogs that act as reverse transcriptase inhibitors (nRTI). These nRTIs, which were initially developed to treat HIV infections, selectively target the reverse transcriptase of the HIV/HBV virus and inhibit its functioning. nRTIs do not cure a patient from HBV, but stagnate levels of virus in the patient (2). Lamivudine is a cytosine like reverse transcriptase inhibitor and is administered to patients suffering from HIV and more recently also to patients infected by HBV. Unfortunately, HBV becomes resistant to Lamivudine after one or two years. Consequently, Lamivudine treatment against chronic HBV becomes ultimately ineffective within a rather short period of time (2). Next to the recurrence of HBV after treatment, there is an economic aspect: treatment can cost thousands of dollars per year and is not available to most patients in developing countries (23).

1.3.2 The Hepatitis B virus and its replication cycle

The hepatitis B virion consists of an icosahedral nucleocapsid with a diameter of ~33 nm, which holds the DNA of the virus. The virion has been studied with cryo-electron microscopy by Watts and coworkers (4). The structure of the hepatitis B capsid they derived is displayed in Figure 1.2. Three different types of protein spikes are embedded in the outer lipid envelope, they are essential for recognition and entry of a host cell. Entry of the virus into a cell probably occurs via endocytosis, but details on how the nucleocapsid enters the cytoplasm are not yet well defined (32; 33).

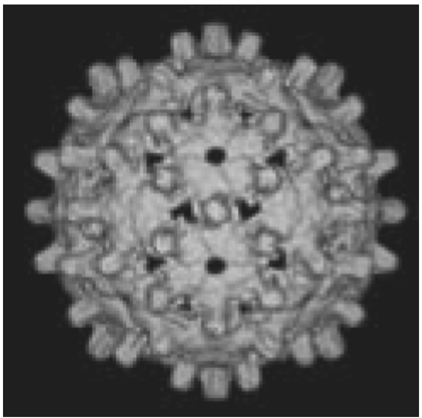


Figure 1.2 : A cryo-TEM image of the nucleocapsid of the hepatitis B virus. (4)

When HBV infects a cell, it loses its viral envelope. The viral core continues its journey to the nucleus and releases its DNA into the nucleus. After which the DNA is converted into covalently closed circular DNA (cccDNA) (34). The viral cccDNA is then transcribed by the RNA polymerase II of the host cell to pregenomic RNA (pgRNA), which is in turn transported to the cytoplasm. Here, protein P and capsid proteins are translated from

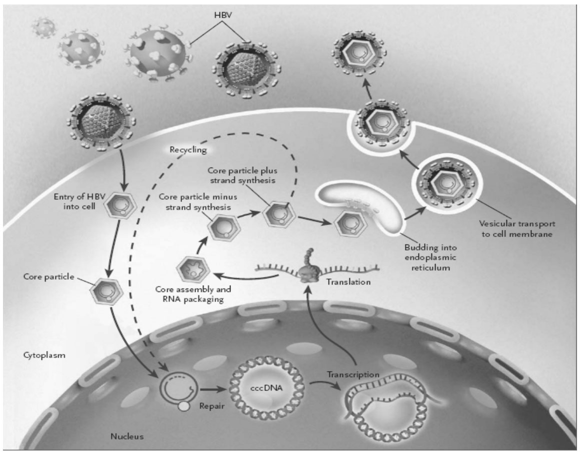


Figure 1.3: The replication cycle of hepatitis B virus (2)

the pgRNA on the ribosome of the host cell. The multi-domain protein P consists of an N-terminal domain (TP), a spacer, a reverse transcriptase domain (RT), and an RNAse-H domain (RH) (6; 33). Binding of the RT-domain to the apical stem-loop of the ε element of the pgRNA (Figure 1.4) (35; 36) initiates encapsidation of the pgRNA-protein P complex into core particles. This ε is a highly conserved RNA element (37) located near the 5'-end of the pgRNA (Figure 1.4). Subsequently, reverse transcription is initiated by synthesis of a 4 nt primer using part of the primer bulge of ε as a template. This 4 nt primer is covalently bound to a tyrosine in the TP domain of the multi-domain protein P (6; 33). After primer synthesis, protein P with its covalently bound primer translocates to acceptor direct repeat 1* (DR1*), located proximal to the 3'-end of the pgRNA. From DR1*, (-)-DNA strand synthesis starts using the pgRNA as template. The (-)-DNA strand serves as a template for the replication of the complementary (+)-DNA strand. Once genome replication is completed, the virus is budded by the Golgi apparatus and exocytosis takes place, resulting in the birth of a new virus particle. Thus, protein P – epsilon interaction is crucial for the viral replication.

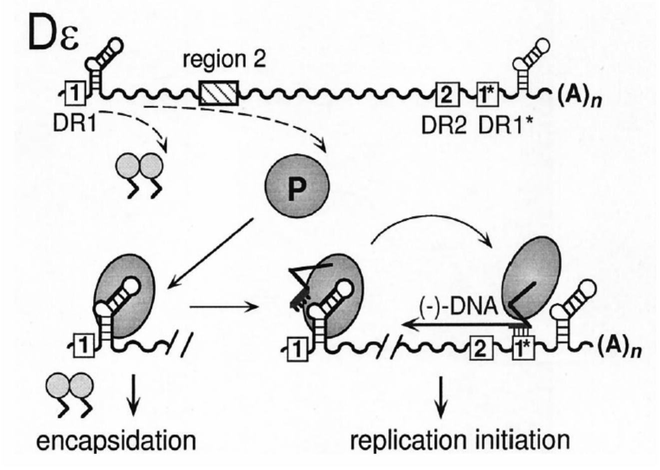


Figure 1.4: Mechanism of replication via the ε element. (5)

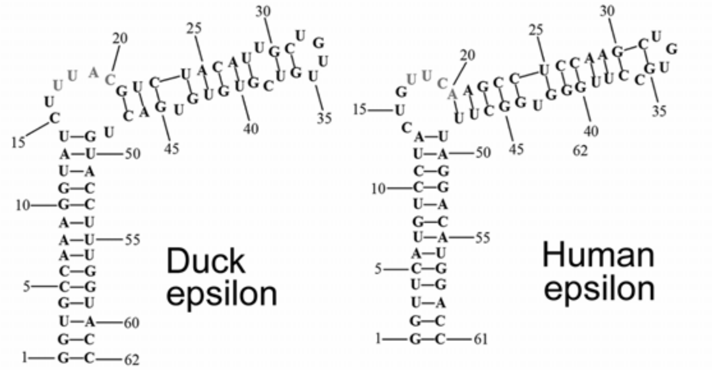


Figure 1.5: Secondary structure of epsilon of duck (left) and human

The secondary structures of ε elements of the pregenomic RNA of HBV of human and duck are shown in Figure 1.5. They each consist of two structure elements, namely the top stem and loop called: apical loop (AL) and the bottom stem and bulge area, called the primer bulge (PB). Both structure elements play a crucial part in the binding and priming of the protein complex (38; 39), and thus ε plays an important role in the replication mechanism of HBV. The most important differences between the ε of human and duck HBV are insertion of a uridine residue in the primer bulge at position 47, the difference in primer sequence, which for duck is UUAC, and for human UUCA, and the different apical loops, in duck a UGUU tetraloop and in human a pseudo-triloop with sequence G-CUGUGC-C. The differences and similarities between the two will be discussed in more detail in chapter 4 of this thesis.

For duck, a fully functional in-vitro cell-free and chaperone-dependent reconstituted P/ε system was developed (5; 19; 33; 36; 38; 40-42). Many of the molecular details of the RT- ε interaction were derived from this in-vitro system. Only recently, a partly functional (only binding) in-vitro system was developed for human HBV (43; 44).

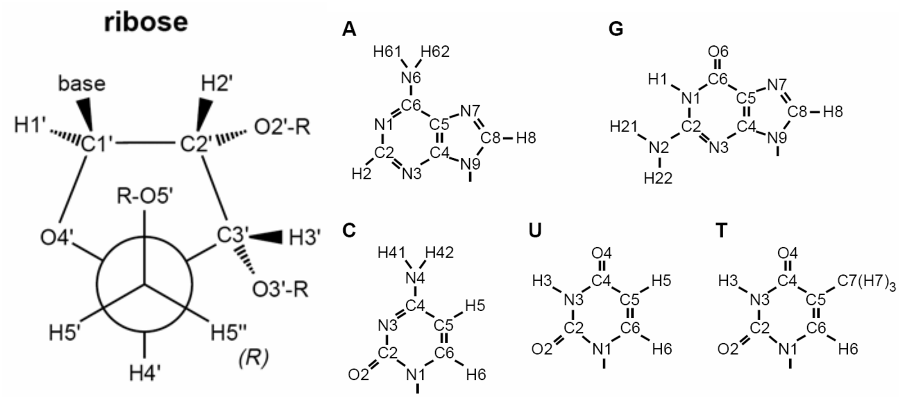


Figure 1.6: Building blocks of RNA molecules: Nucleosides (right), and ribose (left). A Newman projection is made along the C5’-C4’ bond, to indicate the stereochemistry of H5’ and H5’’. Nucleosides commonly found in RNA are Adenine (A), Guanine (G), Cytosine (C), and Uridine (U). Thymidine (T) occurs in DNA instead of Uridine (U), and is not a common base in RNA chains.

1.4 RNA structure

To study RNA structures and their dynamics at atomic detail requires clear definitions and knowledge of the basic building blocks of RNA molecules. Definitions of these building blocks will be described together with a way to describe helix nucleic acid structure. Definitions of the building blocks of nucleic acids can also be found in Saenger (45)

1.4.1 General description of RNA and its building blocks

The building blocks of RNA consist of a ribose sugar (D-furanose) and an attached base (Figure 1.6). The combination of a ribose and a base is called a nucleoside, a combination of a ribose with phosphate group attached on the O5'-atom, and a base is called a nucleotide.

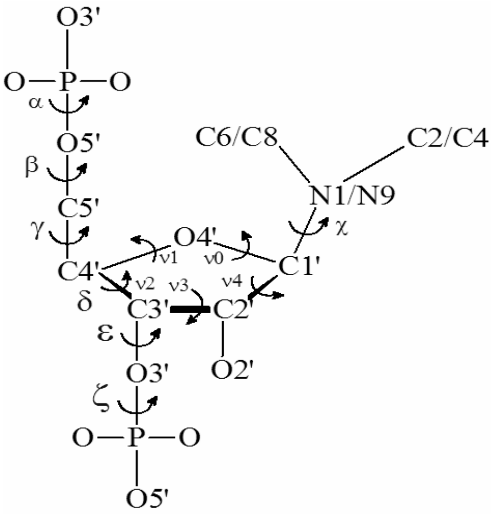


Figure 1.7: All dihedral angles that can vary within RNA.

RNA molecules are oligomers of nucleotides in which sequential ribose sugars are linked via a phosphorus atom P to O3' of the next residue. The backbone is defined by sequential riboses connected to each other via phosphate groups, like shown in Figure 1.7.

In an RNA chain, a large number of dihedral angles can be identified when compared to proteins. The typical dihedral angles in RNA are displayed with their proper names (46) in Figure 1.7. Dihedral angles ν_0 to ν_4, defining the sugar conformation, can be contracted into the pseudo rotation angle *P* (3) given in equation (1.1) .

$$P = \arctan\left(\frac{\left(\nu_4+\nu_1\right)-\left(\nu_3+\nu_0\right)}{2\nu_2\left(\sin\left(36^o\right)+\sin\left(72^o\right)\right)}\right) \tag{1.1}$$

Pseudo rotation angle *P* can be displayed in a pseudo-rotation wheel (Figure 1.8), that provides an overview of allowed ribose-sugar conformations. Bold arrows in this wheel indicate the most preferential sugar puckers, i.e. the N-pucker mode (north) that occurs most in RNA A-helices, and S-pucker (south) mode that occurs mainly in B-helices of DNA.

In RNA and DNA, the sequence of nucleotides is called the primary structure (47). The secondary structure is the sum of the canonical (Watson-Crick and G:U) base pairs, which gives rise to formation of A-helix like structures in RNA. Two examples of RNA secondary structure are shown in Figure 1.5. The tertiary structure is the three-dimensional arrangement of atoms.

Apart from the double-stranded A-helices (48), in some cases triple-stranded helical structures (49) or even quadruple-stranded structures (50) can be formed. In addition, internal bulges, hairpin structures A-platforms etc. are RNA structure elements that have been observed. For instance, the ribosome contains a large variety of RNA structure elements (14). Apart form the common Watson-Crick A:U and C:G basepairs, a variety of alternative basepairs have been observed (51; 52), such as C:C, A:A, G:G, U:C.

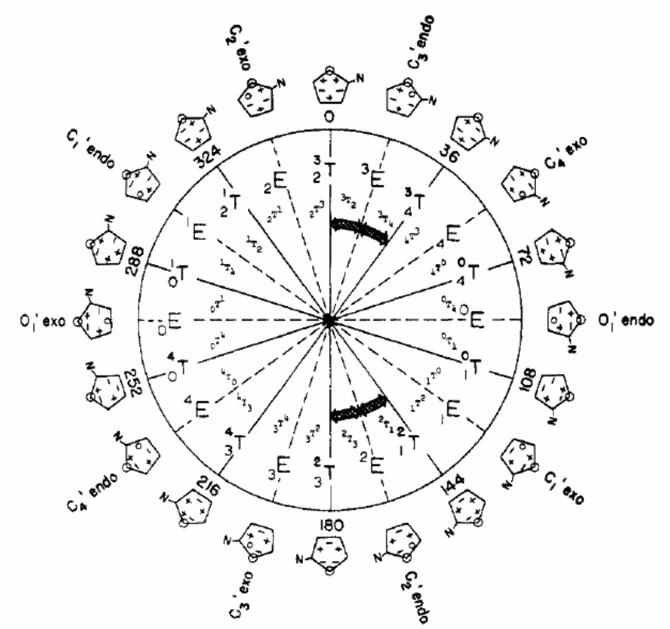


Figure 1.8: The pseudo-rotation wheel provides all possible sugar puckering modes (3)

In general, folding of RNA is hierarchical in that secondary structure is more stable than tertiary structure (47; 53-59). Furthermore, secondary structure forms on shorter timescales than tertiary structure (53; 59). Therefore, secondary structure can be determined largely independently of tertiary structure. Many programs exist (54; 60-63) that predict RNA secondary structure from their nucleotide sequence (primary structure).

The structure features are of major importance for the biological functioning of RNAs, such as information transfer, coding and catalytic activities. The formation of intricate three-dimensional (3D) structures is required for the proper functioning of RNAs (64), therefore studying 3D structures is essential. Although prediction of secondary structure is common practice, prediction of 3D-RNA structures has not been possible so far.

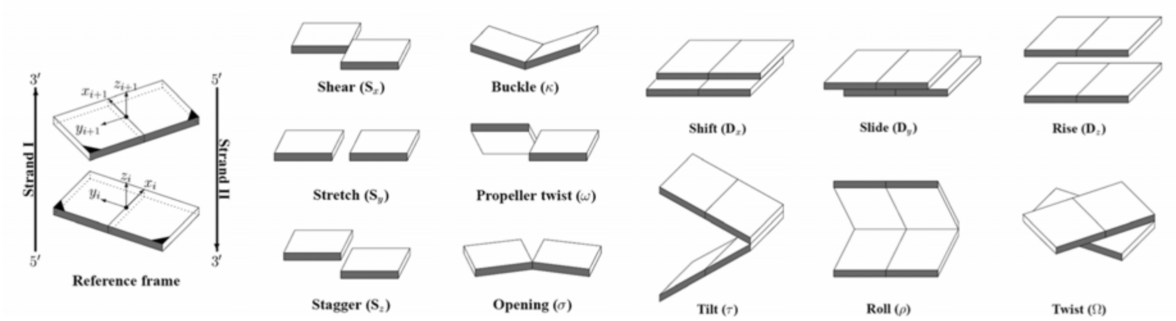


Figure 1.9: Basepair parameters (left) and basepair step parameters (right). Pictures from X3DNA manual (9; 10)

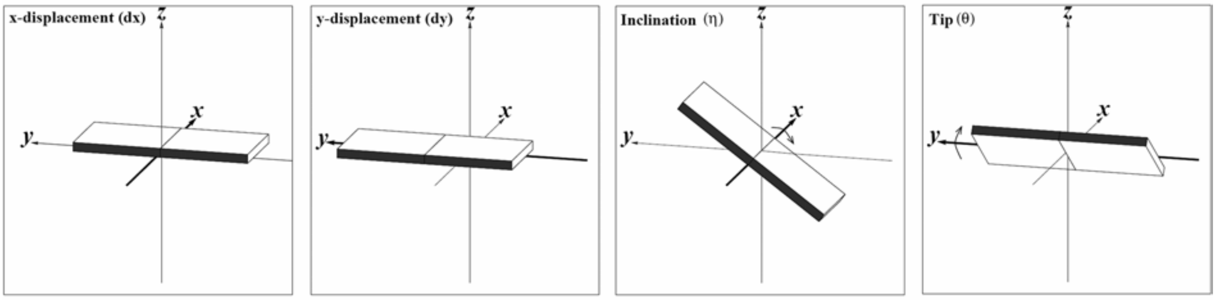


Figure 1.10: Helical parameters. Picture from X3DNA manual (9; 10)

1.4.2 Helical structure elements

Although many alternative RNA secondary folds exist, the majority of nucleotides in RNA participate in standard Watson-Crick base pairing and A-helix formation (65). In RNA structures that have been elucidated so far, close packing of these double-stranded helices is the governing principle of the basic architecture of all higher-order RNA folds (65). These complex three-dimensional architectures in RNA folds (66; 67) hold specific interaction sites that are grafted into the helices and comprise structurally conserved modules that have been found in the three-dimensional architectures of many different RNA molecules. In these specific interaction sites a variability exists, such as variations in Watson-Crick base pairing (65), i.e. mismatches, interaction of triples and quadruples of bases, platforms with pairing between consecutive bases within one strand, bulged-out residues, alternate cross-stacking between bases in different strands, i.e. “interdigitation”, recurring hydrogen-bonding pattern between riboses of consecutive nucleotides in two strands, i.e. “ribose zipper”.

Description of the helix conformation of nucleic acids has not always been a straightforward matter. When the first right-handed double-stranded DNA helix was solved by X-ray diffraction (68), it was evident that a quantitative framework was needed to describe the helix structure and the variations within the helix. The initially chosen helix parameters were recognized as useful, but it became clear that a lack of uniformity existed in the definitions employed. This led to helix parameter values that depended on the nucleotide sequence. In 1988, research groups agreed on symbols and sign-conventions for a complete set of geometric helix parameters (Figures 9 and 10), resulting in what became known as the "Cambridge Accord" (69; 70). The "Cambridge Accord" did not explicitly propose a calculation scheme for these helix parameters. Over time, the need for a systematic calculation scheme became more and more evident (69). Such a systematic calculation scheme has been devised by Babcock et al. (71; 72). The mathematics by Babcock, coupled to a common reference frame for nucleotides (73), allows to calculate helix parameter values in an unbiased fashion, i.e. helix parameters are derived independent of nucleotide sequence. The program 3DNA (9; 10) can be employed to derive such unbiased helix parameters.

1.5 Structure determination of nucleic acids by NMR

If we consider the contents of the online protein databank, X-ray crystallography and NMR spectroscopy are the most important methods to solve three-dimensional nucleic acid structures. X-ray crystallography is often not a feasible method, since growing crystals from nucleic acids is difficult (74-76).

In contrast, in NMR spectroscopy, molecules can be studied in a solution that mimics their natural environment. Recent technological advances like, for example, cryo-probe technology (77) has led to an increased sensitivity and therefore lower concentrations of molecules can be measured. Technology of isotope labeling nucleic acids also reduced the concentration of required sample and increased the amount of information that can be obtained from a sample. Although isotope labeling is beyond the scope of this thesis, some examples will be described for completeness.

Uniform isotope labeling of RNA started in the 90's (78). Later, residue selective labeling using ^{13}C, ^{15}N labeling in combination with ^{2}H labeling was introduced (79-83). Recently, segmental labeling was introduced for both RNA and DNA (84; 85), this allows one to apply the previously mentioned labeling schemes to a selected part of a nucleic acid molecule. Isotope labeling allows one to record heteronuclear correlation spectra, which can assist in assignment of the chemical shifts of heteronuclei and obtaining restraints of the nuclei concerned.

From NMR data, structure information can be extracted. More information will finally lead to a more accurate and precise description of the structure. Commonly used NMR structure information is: NOE distance restraints (86; 87) from which the distance between two nuclei can be derived, J-couplings (88) that give information on dihedral angles. In addition, there are anisotropic restraints that give information on relative orientations, for example residual dipolar couplings (89-91) give information on relative orientation of dipole-dipole vectors, and residual chemical shift anisotropies (92-94) give information on relative orientation of the principal components of chemical shielding tensors respectively.

Another restraint that is not commonly used, is the isotropic chemical shift. Several methods have been proposed and demonstrated that employ chemical shifts as experimental restraints for protein structure determination (95-100). Although chemical shifts have been used to validate nucleic acid structures, chemical shifts alone have never been used as a source of structure information for deriving nucleic acid structure. In chapters 5 and 6 of this thesis, we will show how proton chemical shifts can be applied to determine nucleic acid helix structures.

When all restraints are collected, a structure is calculated using a restrained molecular dynamics protocol(101). In the cases mentioned in this thesis, the program used for this purpose was Xplor v3.851 by Brunger (102). For all types of restraints, energy-potentials are defined, and the topology of the molecule is described in its parameters, for example the parameters given by Clowney (103) and Gelbin (104). An example of a structure calculation protocol: initially the molecule is heated *in silico* to high temperature, e.g. 20000K. Subsequently, the molecule is slowly cooled down to room temperature in a process called annealing (105-109). The high temperatures prevent the structure from getting stuck in a local minimum, to finally obtain a structure fulfilling all structure restraints.

1.5.1 Anisotropic interactions

Anisotropic interactions in liquid state NMR spectroscopy are normally averaged out since molecular orientation is uniformly distributed. Upon addition of an aligning medium, the molecular orientation in solution is no longer uniform. This results in a preferred alignment direction, i.e. for dipolar couplings to be useful in solution NMR studies of macromolecules, the orientation of the molecule must be perturbed such that it becomes slightly anisotropic (110). This makes it possible to observe residual anisotropic interactions and simultaneously conserve small NMR line widths. In contrast, in solid state NMR lines are severely broadened because of the presence of a large amount of relatively strong dipolar interactions in the absence of tumbling. In liquid state NMR, the direction and the degree of alignment are expressed in a tensor: the alignment tensor.

One way to align molecules in the magnetic field is by addition of a small amount of pf1-phages (110) to the sample, or by solving the sample in a poly-acrylamide gel (111). Many alignment media have been reviewed by Prestegard et al. (112) The amount of alignment medium added to a sample determines the extent to which residual anisotropic interactions can be observed, in combination with other aspects such as size and shape of the molecule in the sample. Besides addition of alignment media, there are other methods to achieve alignment in NMR samples e.g. magnetic field alignment (89; 90; 113-122) or electric field alignment (123). The latter will not be discussed further in this thesis.

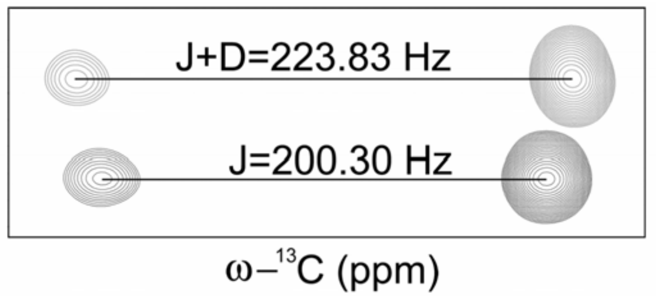


Figure 1.11: An RDC measurement. The bottom lines show a doublet of a C2H2 coupling measured in isotropic conditions, and on top the same doublet in an anisotropic conditions; RDC = 223.83 - 200.30 = 23.53 Hz.

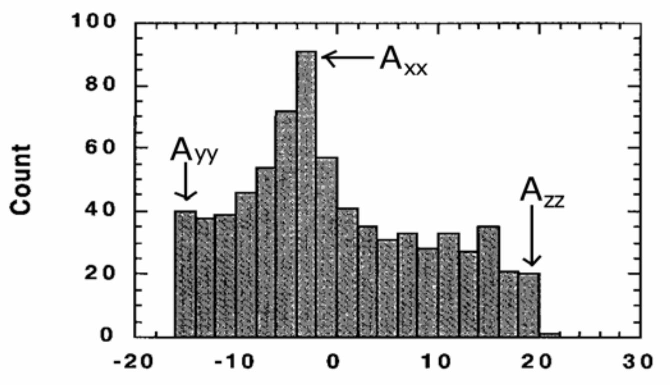


Figure 1.12: Histogram of measured RDCs. Indicated how the principal components of the alignment tensor follow from it (7)

When measuring dipolar couplings in partially aligned samples in liquid state NMR, for example, the largest part of the dipolar interaction is still averaged out and only the residual part of the dipolar coupling is observed, hence the term residual dipolar coupling (RDC). Other observable anisotropic interactions have been described extensively in literature on solid state NMR (124) but are relatively new as restraint in liquid state NMR. Molecular alignment allows us to observe many more residual anisotropic interactions, such as residual chemical shift anisotropy.

For determination of one-bond RDCs, first the J-coupling of the dipole-dipole vectors concerned are measured. The measurements are repeated in an anisotropic phase, i.e. a small amount of an aligning medium is added to the solution. The dipolar coupling is then observed as a change in line splitting, on top of the J-coupling. (89) An example is seen in Figure 1.11, where the lines in the isotropic phase are split up by an amount J; the J-coupling. The lines from the anisotropic phase are split up by an amount J+D for J-coupling together with a RDC contribution.

To extract orientation information from RDC data, it is crucial to define the orientation and the principal components of the alignment tensor with respect to the molecule. A method to obtain the principal components of the alignment tensor has been described by Clore et al. (7) From a histogram of a complete set of observed RDC values, one can directly read the principal components of the alignment tensor as shown in Figure 1.12. Using this method, one assumes that approximately all possible orientations of bond vectors are sampled. For helical nucleic acid structures that are often elongated structures, this is not always the case and other methods have to be used. The advantage of the histogram method is that no structure information is required to determine the principal components of the alignment tensor.

When a preliminary model of the structure is available, the principal components of the alignment tensor together with its correct orientation can be calculated from a preliminary structure. (125) First the gyration tensor of the structure is calculated, which has the same orientation as the alignment tensor. By diagonalizing the gyration tensor and making it traceless, the alignment tensor is defined. Subsequently, the alignment tensor is scaled proportional to the amount of alignment medium that was added to the sample. In this method, it is not assumed that the set of RDCs comprise all possible orientations of bond vectors.

The principal components of the alignment tensor together with its orientation can also be obtained using a grid search (126). A structure from a preliminary structure calculation together with RDC restraints are subjected to a molecular dynamics protocol. In this protocol, a series of RDC structure refinements are started with different axial components (D_a) and the rhombicities (R) for the alignment tensor. During calculations, the alignment tensor is allowed to change its orientation(127). The alignment tensor of the resulting structure that shows the lowest number of violations is chosen, i.e. correct orientation and correct values for D_a and R.

A RDC is observed as a time-averaged coupling, so flexible parts of a molecule will give rise to RDCs that apparently do not match the overall alignment tensor. In other words, the RDC is susceptible to internal dynamics, as shown in section 1.6.3.

Similar to measurement of the RDC, the residual chemical shift anisotropy is measured in both isotropic and anisotropic phases. The resulting difference in the observed resonance frequency between isotropic and anisotropic phase is the residual chemical shift anisotropy, i.e. RCSA.

1.5.1.1 General theory of residual dipolar couplings

Here the equations for residual dipolar couplings in partially aligned molecules starting from the spin Hamiltonian for dipolar interaction will be derived. A line of derivation is followed that closely resembles the one commonly employed in solid-state NMR for deriving the effect of anisotropic dipolar interactions. Derivations of residual dipolar couplings from first principles have been presented previously in different forms (89; 128; 129). The derivations follow in broad lines those presented by Cavanagh et al. (129) and Prestegard et al. (89) As a final motivation, I considered that the equations and formal reasoning presented below could also form a basis for establishing internal dynamics from RDCs, for instance in nucleic acids (see section 1.6.3).

To explain why residual dipolar couplings give rise to a change in line splitting, and how this conveys information on orientation of a dipole-dipole vector, the spin Hamiltonian for dipolar couplings is derived. Let us consider two interacting dipoles I and S with spin ½ in a high magnetic field. This dipolar interaction Hamiltonian can be derived from the classical equation for the interaction energy between two magnetic dipoles (130):

$$E = \frac{\mu_0}{4\pi}\left(\frac{\vec{\boldsymbol{\mu}}_I \bullet \vec{\boldsymbol{\mu}}_S}{r^3} - 3\frac{\left(\vec{\boldsymbol{\mu}}_I \bullet \vec{\mathbf{r}}\right)\left(\vec{\boldsymbol{\mu}}_S \bullet \vec{\mathbf{r}}\right)}{r^5}\right) \tag{1.2}$$

Here, E is the energy in angular frequency units (rad s^{-1}), μ_0 the magnetic permeability of vacuum, $\vec{\boldsymbol{\mu}}_I$ and $\vec{\boldsymbol{\mu}}_S$ are magnetic dipole moments, and $\vec{\mathbf{r}}$ is the vector connecting the two dipoles: the dipole-dipole vector. The quantum mechanical expression for the magnetic moment operator $\hat{\boldsymbol{\mu}}$ for a dipole I is given by:

$$\hat{\boldsymbol{\mu}}_I = \gamma\hbar\hat{\mathbf{I}} \tag{1.3}$$

Here, $\hbar$ is the Dirac constant, γ the gyromagnetic ratio, and $\hat{\mathbf{I}}$ the spin-operator. Substituting this operator (1.3) for $\vec{\boldsymbol{\mu}}_1$ and $\vec{\boldsymbol{\mu}}_2$ in the classical expression for energy of interaction (1.2), we obtain the interaction Hamiltonian for dipolar interaction between spins I and S in angular frequency units (rad s^{-1}) (130):

$$\hat{H} = d\left(\hat{\mathbf{I}}\cdot\hat{\mathbf{S}} - 3\frac{\left(\hat{\mathbf{I}}\cdot\vec{\mathbf{r}}\right)\left(\hat{\mathbf{S}}\cdot\vec{\mathbf{r}}\right)}{r^2}\right) \tag{1.4}$$

With

$$d = \frac{\mu_0\hbar\gamma_I\gamma_S}{4\pi r_{IS}{}^3} \tag{1.5}$$

Here, d is the dipolar interaction energy in angular frequency units (rad s^{-1}). The factor γ_I and γ_S refer to the gyromagnetic ratios for spin I and S respectively in rad s^{-1} T^{-1}, r is the length of the vector connecting the two magnetic dipoles.

This can be rewritten as:

$$\hat{H} = d\left(\hat{I}_x\hat{S}_x\left(1-\frac{3\vec{r}_x{}^2}{r^2}\right)+\hat{I}_x\hat{S}_y\left(\frac{3\vec{r}_x\vec{r}_y}{r^2}\right)+\dots\right) \tag{1.6}$$

The sum in equation (1.6) runs over all possible permutations (x, y, z). This equation can be written in matrix form as:

$$\hat{H} = d\hat{\mathbf{I}}\cdot\mathbf{D}\cdot\hat{\mathbf{S}} \tag{1.7}$$

Here, **D** is the dipolar coupling tensor, which describes the orientation dependence of the interaction of spins I, and S. It can be written in matrix form in terms of its elements in Cartesian coordinates by:

$$\mathbf{D} = \begin{pmatrix} 1-\frac{3r_x{}^2}{r^2} & \frac{3r_x r_y}{r^2} & \frac{3r_x r_z}{r^2} \\ \frac{3r_y r_x}{r^2} & 1-\frac{3r_y{}^2}{r^2} & \frac{3r_y r_z}{r^2} \\ \frac{3r_z r_x}{r^2} & \frac{3r_z r_y}{r^2} & 1-\frac{3r_z{}^2}{r^2} \end{pmatrix} \tag{1.8}$$

To transform this to spherical coordinates, taking θ the rotation around the y-axis and ϕ the rotation around the z-axis, the following transformations can be used:

$$r_x = r\cos(\phi)\sin(\theta) \tag{1.9}$$

$$r_y = r\sin(\phi)\sin(\theta) \tag{1.10}$$

$$r_z = r\cos(\theta) \tag{1.11}$$

Using equations (1.9) to (1.11) and assuming a unit vector ***r***, equation (1.8) can be written as:

$$\mathbf{D} = \begin{pmatrix} 1-3\cos^2(\phi)\sin^2(\theta) & 3\cos(\phi)\sin(\phi)\sin^2(\theta) & 3\cos(\phi)\sin(\theta)\cos(\theta) \\ 3\cos(\phi)\sin(\phi)\sin^2(\theta) & 1-3\sin^2(\phi)\sin^2(\theta) & 3\sin(\phi)\sin(\theta)\cos(\theta) \\ 3\cos(\phi)\sin(\theta)\cos(\theta) & 3\sin(\phi)\sin(\theta)\cos(\theta) & 1-3\cos^2(\theta) \end{pmatrix} \tag{1.12}$$

In the principal axis frame (PAF) of the dipolar coupling tensor, the z-axis of the dipolar coupling tensor is aligned with the vector connecting magnetic dipoles I and S, i.e. *r = (0, 0,* ***r****)*, and therefore the dipole coupling tensor becomes:

$$\mathbf{D}_{PAF} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{1.13}$$

In PAF of the dipolar coupling tensor, the Hamiltonian for dipolar interaction can be written as:

$$\hat{H}_{PAF} = -2d\left(\hat{I}_z\hat{S}_z - \frac{1}{2}\left(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y\right)\right) \tag{1.14}$$

Other non-secular terms resulting from the summation in equation (1.6) can be neglected in the high field approximation. When dipoles I and S are in a high magnetic field, the full spin Hamiltonian for the system is the sum of the Zeeman Hamiltonian $\hat{H}_Z$ and the J-coupling Hamiltonian $\hat{H}_J$ and the dipolar interaction Hamiltonian $\hat{H}_D$:

$$\hat{H} = \hat{H}_Z + \hat{H}_J + \hat{H}_D \tag{1.15}$$

This can be written more explicitly in the PAF as:

$$\hat{H}_{PAF} = \gamma_I\left(1-\sigma_I{}^{iso}\right)B_0\hat{I}_z + \gamma_S\left(1-\sigma_S{}^{iso}\right)B_0\hat{S}_z + 2\pi J\hat{I}_z\hat{S}_z - 2d\left(\hat{I}_z\hat{S}_z - \frac{1}{2}\left(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y\right)\right) \tag{1.16}$$

Here, γ_I and γ_S represent the gyromagnetic ratios (rad s^{-1} T^{-1}), σ_I^{iso} and σ_S^{iso} the dimensionless trace of the chemical shielding tensors divided by three, B_0 is the applied magnetic field (T) and J is the J-coupling constant (Hz).

For a heteronuclear spin system (spin ½), only the terms that commute with the Zeeman Hamiltonian need to be included. The terms $\hat{I}_x\hat{S}_x$ and $\hat{I}_y\hat{S}_y$ in the dipolar Hamiltonian can be dropped, since Zeeman interaction causes fast fluctuations in their contributions making their contribution negligible (to first order). The truncated spin Hamiltonian in the PAF is then:

$$\hat{H}_{PAF} = \gamma_I\left(1-\sigma_I{}^{iso}\right)B_0\hat{I}_z + \gamma_S\left(1-\sigma_S{}^{iso}\right)B_0\hat{S}_z + 2\left(\pi J - d\right)\hat{I}_z\hat{S}_z \tag{1.17}$$

In equation (1.17) it can be seen, that both dipolar interaction and J-coupling have the same Hamiltonian, with the only difference that one evolves with frequency *J* and the other with frequency *d*. For this reason, both contribute to the splitting of their spectral line.

In liquid-state NMR, the measured RDCs are mostly heteronuclear, therefore I will focus on the heteronuclear dipolar spin Hamiltonian like in equation (1.17). The applied magnetic field defines as the quantization axis (z-axis) for the spin-states. The coordinate frame in which the magnetic field defines the z-axis is called the laboratory frame (LAB). To obtain an analytical expression for the dipolar spin Hamiltonian in the lab frame, it is useful to express $\hat{H}_D^{PAF}$ in a frame attached to the molecule: the molecular axis frame (MAF). This frame is unique for the molecule concerned, and allows one to express any bond orientation in the molecule relative to the orientation of the molecular axis frame. To obtain the expression for $\hat{H}_D^{PAF}$ in the LAB, it is required to transform $\hat{H}_D^{MAF}$ to the LAB frame. The last transformation includes a time-dependent rotation, since the orientation of the molecule (MAF) with respect to the applied magnetic field (related to LAB) changes in time due to rotational diffusion. This rotation is written in two steps: first the time dependent rotation is performed, after which time averaging is applied, indicated by an overbar. All these transformations are Euler rotations of the spatial part of the Hamiltonian. They can be summarized as:

$$\hat{H}_{PAF} \xrightarrow{D_{MAF}(\zeta,\theta,\phi)} \hat{H}_{MAF} \xrightarrow{D_{LAB}(\alpha(t),\beta(t),\gamma(t))} \hat{H}_{LAB}(t) \xrightarrow{\overline{(\alpha(t),\beta(t),\gamma(t))}} \left\langle \hat{H}_{LAB} \right\rangle \tag{1.18}$$

The Hamiltonian is rotated from the PAF to the MAF over Euler angles (ζ,θ,ϕ), where the definitions of the Euler angles are given by Brink and Satchler (131). For an exact description and definitions of the Euler rotations of spherical tensors, we refer to Appendix I at the end of this chapter. The second transformation is an Euler rotation over the angles (α,β,γ). Since these angles are time-dependent, the last step is to time-average the (α,β,γ)-dependent terms in the Hamiltonian.

To perform these rotations, it is convenient to describe the dipolar interaction Hamiltonian in irreducible spherical tensors, i.e. the dipolar Hamiltonian in the PAF equals:

$$\hat{H}_{PAF} = d\rho_{20}\hat{\mathbf{T}}_{\mathbf{20}} \tag{1.19}$$

Here, d is the dipolar interaction energy as shown in equation (1.5) and the individual terms ρ_{20} and $\hat{\mathbf{T}}_{\mathbf{20}}$ (heteronuclear) are given by (Appendix I, (132)):

$$\hat{\mathbf{T}}_{\mathbf{20}} = \sqrt{\frac{2}{3}}\hat{I}_z\hat{S}_z \tag{1.20}$$

$$\rho_{20} = \sqrt{\frac{3}{2}}\left(D_{zz} - \frac{1}{3}Tr\left(\mathbf{D}\right)\right) = \sqrt{\frac{3}{2}}D_{zz} \tag{1.21}$$

To express the Hamiltonian in equation (1.19) in the laboratory frame, two sets of rotations as given in (1.18) need to be carried out. The transformation of ρ_{lm} from the PAF to the MAF can be written as (see Appendix I and (130)):

$$\hat{\hat{R}}\left(\rho_{lm}\right) = \sum_{m'=-l}^{+l} \rho_{lm'}D^{l}_{m'm}\left(\zeta,\theta,\phi\right) \tag{1.22}$$

$\hat{\hat{R}}\left(\rho_{lm}\right)$ is a superoperator, that transforms the irreducible representation of the spherical tensor ρ_{lm}, the index l indicates the rank of the tensor, and index m runs from $-l$ to $+l$ and is called the order of the tensor. The terms $D^{l}_{m'm}(\zeta,\theta,\phi)$ are the individual elements (m',m) from the Wigner rotation matrix (see Appendix I and (131; 133)) and the Euler rotation angles are defined as follows (131): 1) rotate ζ degrees around the z-axis of the original axis frame; 2) rotate θ degrees around the y-axis of the newly obtained axis frame resulting from 1; 3) rotate

ϕ degrees around the z-axis of the axis frame resulting from 2. When a second further (time dependent) rotation is performed from MAF to LAB over Euler angles (*α(t),β(t),γ(t))*one obtains:

$$\hat{\hat{R}}\left(\rho_{\mathrm{ln}}\right) = \sum_{m'=-l}^{+l} \sum_{n'=-l}^{+l} \rho_{lm'} D^{l}_{m'n'}\left(\zeta,\theta,\phi\right) D^{l}_{n'n}\left(\alpha(t),\beta(t),\gamma(t)\right) \tag{1.23}$$

Since dipolar interaction is described by a tensor of rank 2, $\hat{\hat{R}}\left(\rho_{20}\right)_{LAB}$ can be written as:

$$\hat{\hat{R}}\left(\rho_{20}\right)_{LAB} = \sum_{n'=-2}^{+2} \rho_{20} D^{2}_{0n'}\left(\zeta,\theta,\phi\right) D^{2}_{n'0}\left(\alpha(t),\beta(t),\gamma(t)\right) \tag{1.24}$$

Hence, the Hamiltonian from equation (1.19) expressed in the laboratory frame:

$$\hat{H}_{LAB} = \hat{\hat{R}}\left(\rho_{20}\right)_{LAB} \hat{\mathbf{T}}_{20} = \hat{\mathbf{T}}_{20} \sum_{n'=-2}^{+2} \rho_{20} D^{2}_{0n'}\left(\zeta,\theta,\phi\right) D^{2}_{n'0}\left(\alpha(t),\beta(t),\gamma(t)\right) \tag{1.25}$$

This equation makes use of the observation that $\hat{\mathbf{T}}_{\mathbf{20}}$ is much simpler to express in the laboratory frame, but the tensors ρ_{20} have their simplest form in the principal axis system of the interaction. Angles α, β, and γ represent the Euler-angles between the molecular axis frame and the laboratory frame. θ is the angle between the z-axis of the principal axis frame and the z-axis of the molecular axis frame and ϕ is the angle between the x-axis of the principal axis frame, and the x-axis of the molecular axis frame, the angle ζ cancels because the element *(0,n')* of the Wigner rotation matrix can be written explicitly as (Appendix I):

$$D^{2}_{0n'}\left(\zeta,\theta,\phi\right) = e^{-i\zeta 0} d^{2}_{0n'} e^{i\phi n'} \tag{1.26}$$

The elements of the Wigner rotation matrix from the MAF to LAB rotation, $D^{2}_{n'0}\left(\overline{\alpha(t),\beta(t),\gamma(t)}\right)$, can be written in terms of normalized spherical harmonics to calculate the effect of time averaging under the influence of a non-uniform probability distribution function(133):

$$D^{2}_{n'0}\left(\overline{\alpha(t),\beta(t),\gamma(t)}\right) = \sqrt{\frac{4\cdot\pi}{5}} \cdot Y^{*}_{2n'}\left(\overline{\beta(t),\alpha(t)}\right) \tag{1.27}$$

The orientation of the molecular axis frame in a magnetic field can be described by the orientation probability density function, which can be expanded in spherical harmonics (89; 134):

$$P\left(\overline{\beta(t),\alpha(t)}\right)=\frac{1}{4\pi}+\sum_{m=-2}^{+2}c_{2m}\cdot\sqrt{\frac{4\cdot\pi}{5}}Y_{2m}\left(\overline{\beta(t),\alpha(t)}\right) \tag{1.28}$$

When a preferred molecular orientation is absent, the values of coefficients c_{2m} drop to zero. By multiplying equation (1.28) with equation (1.27) performing integration over all orientations, whilst yields the time-averaged spherical tensor. The full rotation from MAF to LAB can be written as (89; 134):

$$D^2{}_{n'0}\left(\overline{\alpha(t),\beta(t),\gamma(t)}\right)=\int_0^{2\cdot\pi}\int_0^{\pi}P\left(\overline{\beta(t),\alpha(t)}\right)D^2_{n'0}\left(\alpha(t),\beta(t),\gamma(t)\right)\sin\left(\beta(t)\right)d\beta d\alpha \tag{1.29}$$

After substitution of equations (1.27) and (1.28) in equation (1.29) and applying orthonormality relations between spherical harmonics, this equals:

$$D^2{}_{n'0}\left(\overline{\alpha(t),\beta(t),\gamma(t)}\right)=c_{2n'}\,\delta_{mn'} \tag{1.30}$$

In equation (1.30) $\delta_{mn'}$ is the Dirac delta function, with index m the order of the spherical harmonic function to be rotated (as in equation (1.28)), and $c_{2n'}$ is a constant representing an irreducible part of a tensor expressing the orientation of the MAF in the LAB (134): the Sandle-Williams alignment tensor **A** (135). From now on, the Sandle-Williams alignment tensor will be referred to as the alignment tensor. Thus, after the second rotation combined with time averaging $\hat{\hat{R}}\left(\rho_{20}\right)_{LAB}$ becomes:

$$\hat{\hat{R}}\left(\rho_{20}\right)_{LAB}=\sum_{n'=-2}^{+2}\rho_{20}D^2_{0n'}\left(\zeta,\theta,\phi\right)c_{2n'} \tag{1.31}$$

Hence, the Hamiltonian from equation (1.19) expressed in the laboratory frame after substitution of all Wigner rotation matrix elements (see Appendix I), equations (1.21) and (1.20) in equation (1.31), equals:

$$\begin{aligned}\hat{H}_{LAB}=&(c_{2-2}e^{2i\phi}\sqrt{\frac{3}{8}}\sin^2\left(\theta\right)+c_{2-1}e^{i\phi}\sqrt{\frac{3}{8}}\sin\left(2\theta\right)+c_{20}\frac{3\cos^2\left(\theta\right)-1}{2}\\&+c_{21}e^{-i\phi}\sqrt{\frac{3}{8}}\sin\left(2\theta\right)+c_{22}e^{-2i\phi}\sqrt{\frac{3}{8}}\sin^2\left(\theta\right))\sqrt{\frac{3}{2}}\left(D_{zz}-\frac{1}{3}Tr\left(\mathbf{D}\right)\right)\sqrt{\frac{2}{3}}2\hat{I}_z\hat{S}_z\end{aligned} \tag{1.32}$$

Here c_{20} is the irreducible representation of the z-component of the diagonalized alignment tensor **A**, and $c_{22}+c_{2-2}$ the irreducible representation of a linear combination of x-y of the alignment tensor, and the coefficients c_{21} and c_{2-1} represent the irreducible representation of linear combinations of the off-diagonal terms of the diagonalized alignment tensor; see Appendix I and Snyder (134). Finally, the dipolar Hamiltonian can be written as:

$$\hat{H}_{LAB} = \left(c_{20} \frac{3\cos^2(\theta) - 1}{2} + \frac{c_{2-2} + c_{22}}{2} \sqrt{\frac{3}{2}} \sin^2(\theta)\cos(2\phi) \right) D_{zz} \hat{I}_z \hat{S}_z \tag{1.33}$$

Which can be rearranged to:

$$\hat{H}_{LAB} = \frac{\Delta}{2} \left[\frac{3A_{zz}}{2} \left(3\cos^2(\theta) - 1 \right) + \frac{3}{2} \left(A_{xx} - A_{yy} \right) \sin^2(\theta)\cos(2\phi) \right] 2\hat{I}_z \hat{S}_z \tag{1.34}$$

The angles (ζ, θ, ϕ) are the Euler rotation angles, about which the dipolar vector is orientated with respect to the alignment frame. Δ is the anisotropy of the dipolar interaction tensor defined as $\Delta = D_{zz} - Tr(\mathbf{D}) = D_{zz}$. **A** with its subscripts are the principal components of the alignment tensor (135). For a diagonalized alignment tensor **A**, by definition $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$, furthermore:

$$\mathbf{A} = \begin{pmatrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{pmatrix}$$

$$A_a = \frac{3A_{zz}}{2} \qquad A_r = A_{xx} - A_{yy} \tag{1.35}$$

$$R = \frac{A_r}{A_a} \qquad D_a = d\,\frac{A_a}{2} = d\,\frac{3A_{zz}}{4}$$

Substituting equation (1.35) in equation (1.34) yields:

$$\hat{H}_{LAB} = d\,\frac{A_a}{2} \left[\left(3\cos^2(\theta) - 1 \right) + \frac{3A_r}{2A_a} \sin^2(\theta)\cos(2\phi) \right] \mathbf{2\hat{I}_z \hat{S}_z} \tag{1.36}$$

This can also be written as:

$$\hat{H}_{LAB} = D_a \left[\left(3\cos^2(\theta) - 1 \right) + \frac{3}{2} R \sin^2(\theta)\cos(2\phi) \right] \mathbf{2\hat{I}_z \hat{S}_z} \tag{1.37}$$

Finally, the expression for dipolar coupling becomes (136):

$$D_{IS}(\theta, \phi) = D_a \left(3\cos^2(\theta) - 1 + \frac{3}{2} R \sin^2(\theta)\cos(2\phi) \right) \tag{1.38}$$

Given D_a and R, and the coupling, one may calculate possibilities of θ and ϕ. This together with other structural restraints gives two possible orientations of the bond vector, due to inversion symmetry of the dipolar coupling tensor.

1.5.1.2 Residual chemical shift anisotropy

The residual chemical shift anisotropy (RCSA) of partially aligned molecules can be derived in a similar manner as the residual dipolar couplings (RDCs). After some tedious mathematics, this leads to the equation:

$$\begin{aligned} RCSA(\theta,\phi,\zeta) = \gamma B_0 A_a [\left(\sigma_{zz} - \frac{1}{3} tr\boldsymbol{\sigma}\right)\left(3\cos^2\theta - 1\right) + \left(\sigma_{xx} - \sigma_{yy}\right)\cos\left(2\zeta\right)\sin^2\theta \\ + \frac{1}{8}\frac{A_r}{A_a}\left(\sigma_{xx} - \sigma_{yy}\right)\left(\cos\left(2\zeta\right)\cos\left(2\phi\right)\left(3 + \cos\left(2\theta\right)\right) + 4\sin\left(2\zeta\right)\sin\left(2\phi\right)\cos\theta\right)] \end{aligned} \tag{1.39}$$

Here, σ_{xx}, σ_{yy} and σ_{zz} represent the principal components of the chemical shielding tensor. θ,ϕ, and ζ are the Euler-angles as defined by Brink and Satchler (131) of the chemical shielding tensor and expressed in the molecular axis frame. As shown in equation (1.39) RCSAs convey the same information as RDCs do, and therefore they can be employed in principal like RDCs as structure restraints. In solid-state NMR, chemical shift anisotropy has a long-standing history as useful structural parameter (124). Residual chemical shift anisotropies (RCSA) derived for partially aligned molecules have only recently been introduced as alternative or additional global restraints to RDCs in structure calculation of proteins and nucleic acids (92-94; 137).

For measurement of RCSAs, the resonance position of a spin needs to be determined in both the presence and in the absence of an alignment medium to measure the RCSA. To measure RDCs, the line splitting on a doublet needs to be measured in the presence and absence of alignment but for RCSAs only the resonance frequency is of importance, i.e. one observes singlet lines. Consequently, the signal-to-noise is a factor of two higher in RCSA NMR spectra as compared to NMR spectra required for measuring RDCs, while the number of resonances is reduced by a factor of two leading to less overlap. Some disadvantages are also evident. For instance, the orientation of the anisotropy vector of the chemical shift tensor relative to the local frame is not always known. Furthermore, the chemical shift tensor may change shape and orientation upon a change in molecular conformation. The major problem is the sensitivity of the chemical shift (reference) to changes in structure and environment, such as temperature changes, and addition of an alignment medium. This makes it difficult to produce and measure on a sample in both isotropic and anisotropic conditions.

This problem has been addressed for a DNA with a helical conformation (138). They observed that the principal components of the ^{13}C chemical shift tensors of C1’ to C5’ in the ribose of RNAs have the largest shielding along the C-O bond. They showed that density functional theoretical calculations (139) could predict the orientation of the chemical shielding tensor. The orientation of the alignment tensor was known from RDC measurements. The orientation of the chemical shielding tensor with respect to the alignment tensor can now be calculated, providing additional orientation restraints.

1.5.1.3 Orientation probability density function in magnetic field alignment

The derivation for residual dipolar couplings given in section 1.5.1.1 holds for alignment with additional agents that electrostatically align molecules in the magnetic field. However, partial alignment can also be conferred by the intrinsic properties of a molecule, such as the anisotropy of the magnetic susceptibility tensor χ. Nucleic acids have a high degree of aromaticity; this induces a partial alignment of a molecule in strong homogeneous magnetic fields and thereby gives rise to magnetic field-induced RDCs (mRDCs) (89; 91; 113; 114; 121; 122).

For magnetic field alignment, a similar derivation can be made as described in the section 1.5.1.1. The anisotropy of the magnetic susceptibility tensor determines the degree of alignment. The orientation probability density function of the magnetic susceptibility tensor χ with respect to the main field of the magnet can be written as (91):

$$P(\alpha,\beta)=\frac{1}{4\pi}\left(1+\frac{B_0^2}{4k_BT}\left(\chi_a\frac{8}{3}\sqrt{\frac{\pi}{5}}Y_{20}(\alpha,\beta)+\chi_r2\sqrt{\frac{2\pi}{15}}\left(Y_{2-2}(\beta,\gamma)+Y_{22}(\alpha,\beta)\right)\right)\right) \tag{1.40}$$

Following the same treatment as equations (1.29) to (1.38), one can derive (121; 129):

$$D_{IS}(\theta,\phi)=-\frac{\mu_0}{4\pi}\frac{\hbar\gamma_I\gamma_S}{\pi r_{IS}^3}\frac{B_0^2}{15\mu_0k_BT}\left(\frac{1}{2}\Delta\chi\left(3\cos^2(\theta)-1\right)+\frac{3}{4}\delta\chi\sin^2(\theta)\cos(2\phi)\right) \tag{1.41}$$

Here:

$$\Delta\chi=\chi_{zz}-\frac{\left(\chi_{xx}+\chi_{yy}\right)}{2} \tag{1.42}$$

$$\delta\chi=\chi_{xx}-\chi_{yy} \tag{1.43}$$

Equation (1.41) is comparable to equation (1.38) to describe alignment in alignment media.

The splitting due to scalar coupling is independent of the field strength, whereas the mRDC is quadratically dependent on the strength of the magnetic field, i.e. equation (1.41). For nucleic acids, it was found that the magnetic susceptibility tensor χ_b of a single nucleobase can be calculated with good accuracy and that the molecular magnetic susceptibility tensor can be calculated as the sum of individual magnetic susceptibility tensors of the nucleobases (114). Therefore, the principal components of the alignment tensor can be determined.

Orientation probability functions similar to equation (1.40) have been derived for solutions where steric alignment (140) results in a preferred orientation of a molecule in solution; and also for solutions where electrostatic alignment (125), where electrostatic repulsion results in alignment of the molecule in solution.

1.5.2 Isotropic Chemical shifts

Chemical shifts provide detailed information on the structure and electronic properties of biological molecules in solution, crystalline, and non-crystalline states (96-100; 141-143). The importance of chemical shifts as carrier of structure information is highlighted by their increasing prominence in the field of protein structure biology. They are commonly employed to establish the protein secondary structure and backbone torsion angles in polypeptides (141; 144). Recently, it has even been demonstrated that the three-dimensional structure of a protein can be derived, using chemical shifts as sole source of experimental structure parameters (97-100; 142). Such applications make use of empirical structure-chemical shift relationships that have been established from analyses of large databases of protein structures and their related chemical shift (141; 142). Density functional theoretical calculations (139) provided insight in the physical basis for these relationships between structure and chemical shift (96; 143; 145). Structural information contained in chemical shifts is different in nature from that provided by other restraints, since they provide unequivocal information about the relative spatial locations or orientations of different residues in a biomolecular sequence (100; 146). In contrast, the chemical shift associated with a specific atom is a summation of many contributing factors (147-149). This makes reliable identification of interaction partners in proteins very difficult, even though they may be substantially influenced by contacts between residues, such as proximity to aromatic rings and hydrogen bonding that is at very different locations in the protein sequence. If such effects could be interpreted in depth, they would enable the characterization of the detailed environment of virtually every atom in the protein structure and thus the determination of a unique overall conformation.

However, for nucleic acids, empirical relations between three-dimensional conformation and proton chemical shifts for a nucleotide in a triplet sequence (96; 143; 150; 151) have been derived. Due to the prominence of ring current and magnetic anisotropy effects, the proton conformational chemical shift can be interpreted empirically (96; 143). This means that the calculation of the proton chemical shifts from a given three-dimensional structure based on ring-current effects, magnetic anisotropies, and electric field polarization, is highly accurate; for instance, the RMSDs between predicted and observed proton chemical shift is smaller than 0.08 ppm for an RNA helix and smaller than 0.15 ppm for the more mobile loop and bulges. This accuracy approaches the one obtained from purely empirical relationships; the accuracy of the experimental relation between proton chemical shifts and structure in a helix (150; 151) is circa 0.02 ppm. So far, these (semi)-empirical chemical shift-structure relationships have not been extensively employed to establish structure information.

The *de-novo* derivation of the three-dimensional NMR structure of nucleic acids from chemical shifts has not been described in literature, but will be described in chapters 5 and 6 of this thesis. Relationships have also been established between imino protons chemical shift and the $^1J_{HN}$-coupling across the hydrogen bond (152). This information can be used to derive mRDC-values from measurement at a single magnetic field strength via comparison of the calculated $^1J_{HN}$-coupling in comparison with experimental observed line splitting (153; 154). Furthermore, both the imino proton chemical shift and the $^1J_{HN}$-coupling show an exponentially decaying dependence on the length of the hydrogen bond as evident from DFT calculations (152). Thus, these relationships can be employed in structure derivation, e.g. as additional information/restraints in chemical shift based structure calculations as described in Chapter 5.

Relationships between heteronuclear shifts and conformation have only been established to a limited extent for nucleic acids. Empirical relations between ^{13}C chemical shifts and puckering of deoxyribose (DNA) have been established from solid-state NMR experiments (155). An analysis in our lab of the ^{13}C chemical shifts of ribose carbon atoms in PDB-deposited RNA structures (Fonville et al., manuscript in preparation; personal communication) shows strict mutually linear relationships; these relations can in turn be connected to changes in the sugar puckering as well as to changes in other torsion angles. Furthermore, the dependences of $^{13}C/^{15}N$-chemical shifts of ribose - and base carbons on χ-angle for N- and S-puckered ribose sugar ring states have been established for RNA and DNA via various DFT calculations on either isolated mono-nucleotides or di-nucleotides (156) (Fonville et al., manuscript in preparation). Experimental validation of these dependences has partly been carried out. These dependences could form interesting new and complementary structure restraints in chemical shift based structure derivations of nucleic acids (see for instance Chapter 5).

1.5.2.1 Proton chemical shifts and their calculation

As pointed out above, chemical shift values carry important structural information. The chemical shift of a nuclear spin depends on the surrounding electron density as well as on changes in the local magnetic field induced by ring-currents and/or magnetic anisotropies. The electron density and corresponding chemical shifts can be obtained, in principle, by *ab-initio* quantum mechanical calculations (143; 157-159). In a similar vain, ring-current parameters and parameters for the magnetic anisotropy can be derived from quantum-mechanical calculations (143; 157; 159). Progress in quantum mechanical computational procedures and computer hardware has made it possible to perform such calculations for molecular fragments large enough to reflect the essential features of the local environment (95-100; 141; 142; 160-164).

For chemical shift calculations it is operationally expedient to divide a molecule into a number of fragments, that is, into a fragment A, where the nucleus of interest resides, and that has a conformation defined with respect to a reference conformation, and a number, *n*, of fragments B interacting with A. The chemical shift of spin l in fragment A, δ_l, can then be divided into two categories, namely a conformation-independent part, $\delta_{l,ref}$, and a conformation-dependent part, $\delta_{l,conf}$.

$$\delta_l = \delta_{l,ref} + \delta_{l,conf} \tag{1.44}$$

The conformation-independent part above represents the chemical shift of nucleus l in fragment A in its reference state. The conformation-dependent part can result from two sources. 1) They can result from changes in the local environment, $\delta_{l,lcA}$, that is, changes in fragment A with respect to its reference state, e.g. changes in torsion angle, bond length, bond angle etc. 2) They can stem from changes in the interaction of nucleus l in A with the other molecular fragments B_j, $\delta_{l,j,B}$. Substituting this information in equation (1.44) leads to the following expression.

$$\delta_l = \delta_{l,ref} + \delta_{l,lcA} + \sum_{j=1}^{n} \delta_{l,j,B} \tag{1.45}$$

For ^{1}H nuclei, the conformational chemical shift of nucleus l in A does not depend strongly on the torsion angle changes in A. Consequently, $\delta_{l,ref}$ of the ^{1}H nuclei in, for example, the sugar moieties of nucleic acids can appropriately be defined as belonging to the individual C-H fragments and $\delta_{l,lcA}$ can be taken to be zero. In contrast, for heteronuclei, the chemical shift of a particular nucleus l, may be affected quite strongly by torsion angle changes around this nucleus (163). For instance, for $\delta_{l,ref}$ of a ^{13}C nucleus in ribofuranose, it would be operationally more convenient to use as a reference state the S-puckered conformation of the sugar. The term $\delta_{\mathrm{l,lcA}}$ could then describe the chemical shift changes resulting from deviations from the S-puckered state. The contributions to the chemical shift by fragments B, $\delta_{l,j,B}$ can be split up into a number of terms, namely the ring current contribution δ_{rc}, the magnetic anisotropy contribution δ_{ma}, the electric field contribution δ_E, and finally a so-called charge transfer term δ_{CT}. The chemical shift δ_l of the resonance of nuclear spin l in molecule C can then formally be written as:

$$\delta_l = \delta_{l,ref} + \delta_{l,lcA} + \sum_{j=1}^{n} (\delta_{rc} + \delta_{ma} + \delta_E + \delta_{CT})_{l,j} \tag{1.46}$$

Here, $\delta_{l,lcA}$ can be taken equal to zero for ^{1}H nuclei, as described above. The term δ_{CT} is of relevance only when nucleus *l* can form a hydrogen bond with fragment B, like for an imino proton *l* in a base forming a base pair with base B. Non-exchangeable protons usually do not form hydrogen bonds. Consequently, δ_{CT} can be set to zero. The ring-current contribution, δ_{rc} and the magnetic anisotropy contribution, δ_{ma}, form for nucleic acids the major contributors to the conformation-dependent part of the chemical shift of non-exchangeable protons. The term δ_E can essentially be set to zero for protons, due to their small electron polarizabilities. Finally, it should be noted with regard to the above equation that chemical shifts obtained by means of quantum mechanical calculations can be generally interpreted, or rephrased and parameterized, in terms of the separate terms mentioned in equation (1.46). In the following subsections, each of the main contributors to the conformation-dependent shift is described in more detail.

1.5.2.1.1 Ring current effects

Ring current effects form the dominating contributors to the conformational part of the proton chemical shifts of nucleic acids (96; 143; 164). They arise because electrons are delocalized in an aromatic ring, so that an external magnetic field induces a ring current, according to the Maxwell equations. Thereby, they induce a local magnetic field that opposes the external magnetic field, according to Lenz's law. The contribution of the ring current effect to the chemical shift can be written in a general form (143; 157):

$$\delta_{rc} = IBG(\mathbf{r}) \tag{1.47}$$

Here, *I* has a unit value for benzene and is called the ring-current intensity factor. The constant *B* is an empirical parameter adjusted to a value of 2.13 10^{-6} Å. The constant B is chosen such that for benzene, ring-current shifts δ_{rc} of its own protons equal -1.5 ppm, when the delocalized electrons circulate 0.61 Å above and below the plane of the aromatic ring. The term *G(**r**)* is a geometric factor, with **r** the vector from the observed nucleus to the center of the ring that gives rise to the ring-current. The term *G(**r**)* is orientation dependent and approximately proportional to r^{-3}. Two popular models to calculate *G(**r**)* are those of Johnson and Bovey (165) and of Haigh and Mallion (166-168). The Johnson-Bovey model (165) attributes the shift to current loops above and below the plane of the aromatic ring. The Haigh-Mallion model is based on the Hückel molecular orbital theory (166-168). The term *G(**r**)* resulting from the two approaches is very similar. Pullman and collaborators (159) employed the Johnson–Bovey formulation to parameterize their quantum mechanical chemical shift calculations. Case and coworkers parameterized their quantum-mechanically calculated chemical shifts not only in terms of the Johnson-Bovey model but also in terms of the Haig-Mallion model (143; 157). Wijmenga et al. (96; 143) employed the Johnson-Bovey model parameterizations (157; 165) in their program NUCHEMICS to calculate the ring-current chemical shifts and found that the chemical shift predictions for non-exchangeable proton were highly accurate for both parameterizations, e.g. for RNA helices the RMSD between prediction and observed was ca. 0.08 ppm.

1.5.2.1.2 Local magnetic anisotropy

The local magnetic anisotropic contribution to the chemical shift of nucleus *a* can be formulated as (143; 169):

$$\delta_{ma,a} = \frac{1}{3r^5}\sum_{\alpha,\beta}\left(3r_\alpha r_\beta - r^2\delta_{\alpha\beta}\right)\left(1.967R_{\alpha\beta} - 5.368Q_{\alpha\beta}\right) \quad \vee \quad \alpha,\beta \in \{x,y,z\} \tag{1.48}$$

Here, r is the distance between nucleus a and the center of the local magnetic anisotropy located at atom b; $r_{\alpha\in\{x,y,z\}}$ is the α-component of the vector **r** connecting nucleus a and the center of local magnetic anisotropy at atom b. $\delta_{\alpha\beta}$ is the Dirac delta function. The term $R_{\alpha\beta}$ is the αβ-element of the diamagnetic part and $Q_{\alpha\beta}$ is the αβ-element of the paramagnetic part of the magnetic susceptibility tensor of atom b. To calculate the conformational shift stemming from magnetic anisotropy, values of the parameters $R_{\alpha\beta}$ and $Q_{\alpha\beta}$ need to be set. The parameterization given by Giessner-Prettre (159) is employed in NUCHEMICS (96; 143).

1.5.2.1.3 Electric Field effects and charge transfer effects

The contribution of an electric field to the chemical shift of a nucleus is determined by (143):

$$\delta_E = AE_{//} + BE^2 \tag{1.49}$$

Here, E is the electric field at the nucleus and $E_{//}$ the component of electric field at the nucleus parallel to the internuclear bond vector. Application of different computational methods (170) indicated that A may vary between 2.5 and 3.0 10^{-12} esu. For nucleic acids, the value of 2.9 10^{-12} esu (159) appears appropriate and has been chosen by Wijmenga c.s. in NUCHEMICS (143). For B, Giessner-Prettre and Pullman (159) derived a value of 0.74 10^{-18} esu and this value was used in NUCHEMICS (143). The electric field E at the nucleus of atom H can be calculated using the Coulomb law (143):

$$E = \sum_{j=1}^{N} \frac{q_j \vec{r}_j}{4\pi\varepsilon_0\varepsilon_r r_j^{\,3}} \tag{1.50}$$

Here, j is an index to number all charges present, and r_j is the distance between the nucleus and the charges q_j; ε_0 is the dielectric constant of vacuum and ε_r the relative dielectric constant. The value of relative dielectric constant equals 80 in H_2O, while in and around a nucleic acid molecule this value can be as small as 4. The value of the parameter ε_r depends therefore on the type of intervening space between a charge and the nucleus of interest. Due to the small value of ε_r, the effect of electric fields is negligible for proton chemical shifts in RNA (143). To account for the distance-dependence of the dielectric constant the following equation is employed in NUCHEMICS, $\varepsilon_r(r) = 4r$ is used, where r is the distance in Angstrom (143). This approach is adapted from the one employed in molecular mechanics calculations to mimic solvent effects as well as counter-ion screening.

1.6 Dynamics

NMR is ideally suited for determination of both structure and internal dynamics of biomolecules, i.e. the rotational diffusion rates and internal motion. These dynamics are mostly derived from spin relaxation data, but NOEs, RDCs, PREs (paramagnetic relaxation enhancement) (171-173), and PCSs (pseudo-contact shifts) (174) also form a source for dynamics information. The NMR accessible timescales cover essentially all biologically relevant molecular transitions as shown in Figure 1.13.

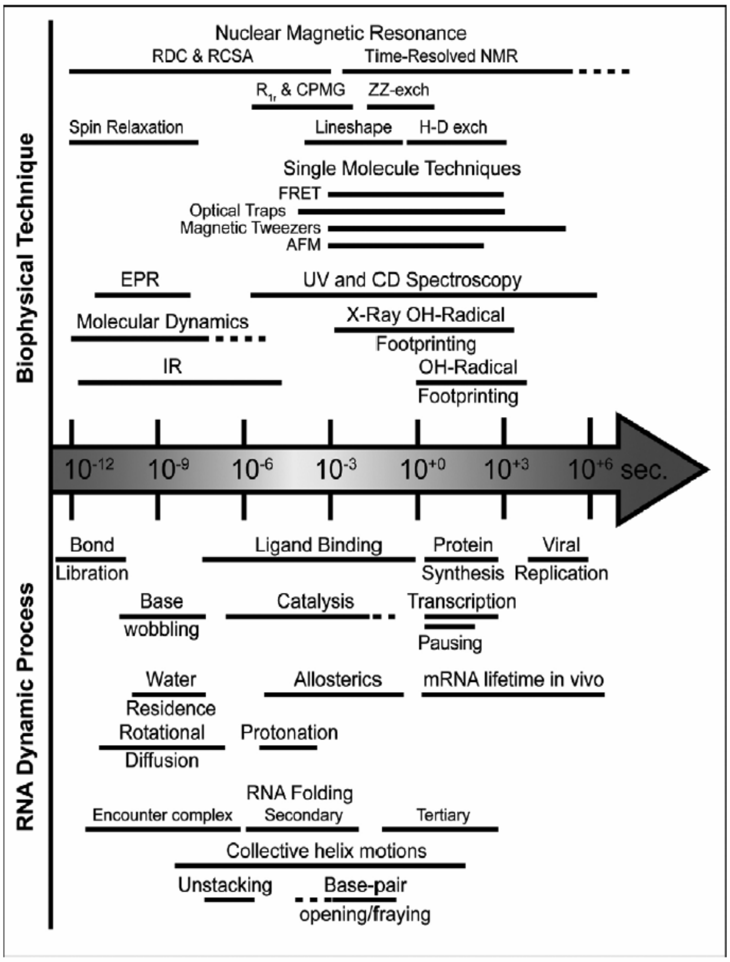


Figure 1.13: Timescales in Biophysical chemistry. (1)

In this section, we first describe how information about dynamics can be obtained from spin relaxation, i.e. NMR spin relaxation measurements. Second, we will describe how information on dynamics can be obtained from residual dipolar coupling information. RDCs have been used, for example, to visualize the motions of an elongated RNA molecule.

1.6.1 NMR Spin Relaxation

For spin relaxation in nucleic acids, two mechanisms of spin interaction are most important (88): 1) the dipole-dipole interaction and 2) the chemical shift anisotropy. 1) In the case of dipole-dipole interaction between two spins, rotational diffusion of a dipole-dipole vector relative to the external magnetic field modulates the interaction. 2) The chemical shift of a nucleus expresses the shielding of the external magnetic field by its environment. This shielding depends on the orientation of the chemical structure relative to the external magnetic field, i.e. it depends on the orientation of the principal axes of the chemical shift tensor relative to the external magnetic field.

When the rates of these motions correspond to nuclear transition frequencies (88; 175-177), rotational diffusion, and internal mobility induce fluctuations in a local magnetic field around a nucleus. These fluctuations may cause transitions in the spin-states of this nucleus (T_1 relaxation and NOE-transfer), and dephasing of coherent magnetization (T_2 relaxation). The Bloch-Wangsness-Redfield theory provides the theoretical framework for relating molecular motion to NMR spin relaxation times (178). How information on molecular dynamics can be extracted from NMR spin relaxation data can be found in a wide array of reviews (88; 175-177; 179-183).

For example, the reciprocal of T_1 NMR relaxation times of spin I due to dipolar interaction with spin S and due to chemical shift anisotropy ($R_{1,I}^{D/CSA}$ and $R_{2,I}^{D/CSA}$) can be expressed as linear combinations of so-called power spectral-density functions *J(ω)* (88; 177):

$$\frac{1}{T_{1,I}^{D}} = A_D^2[\frac{1}{3}J(\omega_I - \omega_S) + J(\omega_I) + 2J(\omega_I + \omega_S)] \qquad (1.51)$$

$$\frac{1}{T_{1,I}^{CSA}} = 4A_{CSA}^2 B_o^2 J(\omega_I) \qquad (1.52)$$

with

$$A_D^2 = \frac{3}{10}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2\gamma_S^2\hbar^2}{r_{IS}^6} \tag{1.53}$$

$$A_{CSA}^2 = \frac{3}{10}\frac{1}{4}\gamma_I^2\delta_z^2\left(1+\frac{\eta^2}{3}\right) \tag{1.54}$$

Here, μ_o is the magnetic susceptibility of vacuum, γ is the gyromagnetic ratio, $\hbar$ is the Dirac constant, r_{IS} is the distance between spin I and S, δ_Z is the z-component of the chemical shift tensor of spin I and η is anisotropy of the chemical shift tensor of spin I.(see appendix). Similar equations exist for T_2 relaxation and NOE-transfer rates.

These functions measure the power distribution of motion-induced frequencies; this distribution depends on the motional model for the molecular system. Below, the most relevant motional models, their corresponding spectral density functions, and the resulting NMR spin relaxation of spins in nucleic acids are described.

The simplest motional model is that of a rigid spherical molecule, for which the spectral density function equals (184):

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{\left(1+\omega^2\tau_c{}^2\right)} \tag{1.55}$$

τ_c is the rotational correlation time of the spherical molecule and ω is the frequency at which the spectral density is considered. Internal motion of the dipole-dipole vector can be accounted for in a model-free manner by the Lipari-Szabo approach as follows (88; 176; 177):

$$J(\tau_c,\tau_e,\omega) = \frac{2}{5}\frac{S^2\tau_c}{(1+\omega^2\tau_c^2)} + \frac{2}{5}\frac{(1-S^2)\tau_e}{(1+\omega^2\tau_e{}^2)} \tag{1.56}$$

with, $\tau_e^{-1} = \tau_c^{-1} + \tau_i^{-1}$. Here, τ_i^{-1} is the reciprocal of the time scale for internal motion. The parameter S^2 is the squared generalized order parameter; it is a measure of the spatial restriction of the internal motion. The allowed range for S^2 is from 0 to 1, with $S^2 = 1$ in the absence of internal motions and $S^2 = 0$ for completely unrestricted isotropic internal motion. Lipari and Szabo (177; 185; 186) showed that internal motions for spin-systems can be described with a single spectral density function, as given in equation (1.56). This is true under the following assumptions: 1) the internal motions are much faster than the overall molecular rotational diffusion $\tau_i >> \tau_c$, 2) the internal (but not overall) motions are in the extreme narrowing limit. In the limiting case of infinitely fast internal motions, the second term in $J(\omega)$ in equation (1.56), may be neglected. The resulting expressions for $J(\omega)$ does not depend on τ_e. In that case, the spectral density function only depends on the generalized order parameter and the overall rotational correlation time, like in equation (1.55). If the internal motions are not in the extreme narrowing limit one can construct successively better approximations for the correlation function of intramolecular motions, consisting of a growing number of exponential terms and thus spectral density functions. In particular, Clore et al. (177; 187) proposed a two-exponential correlation function.

In the presence of an axially symmetric diffusions tensor, that is common in nucleic acids, the spectral density function for a dipole-dipole vector that is rigidly attached to a RNA/DNA is then (88; 177):

$$J(\tau,\omega) = \sum_{j=0}^{2} A_j \frac{2}{5} \frac{\tau_j}{(1+\omega^2\tau_j^2)} \tag{1.57}$$

with

$$A_0 = \frac{1}{4}(3\cos^2\theta_o - 1)^2 \tag{1.58}$$

$$A_1 = \frac{3}{4}\sin^2 2\theta_o; \tag{1.59}$$

$$A_2 = \frac{3}{4}\sin^4\theta_o \tag{1.60}$$

and

$$(\tau_0)^{-1} = 6D_\perp \tag{1.61}$$

$$(\tau_1)^{-1} = 5D_\perp + D_\parallel \tag{1.62}$$

$$(\tau_2)^{-1} = 2D_\perp + 4D_\parallel \tag{1.63}$$

Where, θ_o is the angle of the dipole-dipole vector with the symmetry axis of the axially symmetric diffusion tensor **D**. Thus, the relaxation rate of a spin depends on the orientation of the nuclear interaction vector with respect to the symmetry axis of the diffusion tensor. The time constants τ_j relate to the overall tumbling of the molecule with τ_o (= $(6D_\perp)^{-1}$) representing the tumbling time of the symmetry axis (usually the long axis) of the molecule. For a rigid asymmetric molecule the spectral density function contains 5 terms as described in

Korzhnev et al. (177). Internal motion of the dipolar interaction vector of the spin can be accounted for in a model-free manner by the Lipari-Szabo approach as follows (88; 176; 177):

$$J(\tau,\omega)=\frac{2}{5}\sum_{j=0}^{2}A_j\left[\frac{S^2\tau_j}{(1+\omega^2\tau_j^2)}+\frac{(1-S^2)\tau_{e,j}}{(1+\omega^2\tau_{e,j}^2)}\right] \tag{1.64}$$

Here, A_j and τ_j are as defined above for the rigid axially symmetric rotor. Further, $\tau_{e,j}^{-1}=\tau_j^{-1}+\tau_i^{-1}$ with τ_i the time scale for the internal motion. The parameter S^2 stands again for the squared generalized order parameter. The angle θ_o is the angle of the dipolar interaction vector with the symmetry axis of the axially symmetric diffusion tensor **D**.

It is interesting to consider two other common motional models: wobbling *on* a cone and wobbling *in* a cone. According to the model of wobbling *on* a cone a spin-spin vector rotates *on* a cone with a fixed half-opening angle. In Figure 1.14 motion *on* a cone is depicted; a bond vector rotates around the z-axis of the depicted axis frame with a fixed angle θ_0. In Figure 1.14, wobbling *in* a cone would be described by a spin-spin vector moving around the z-axis with half opening angle anywhere between $-\theta_0$ and $+\theta_0$. For wobbling *on* a cone the order parameter S is given by,

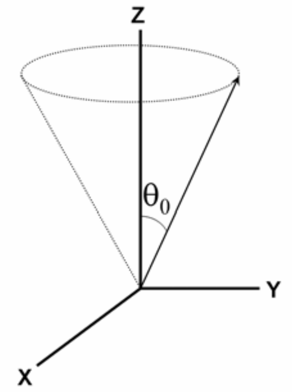


Figure 1.14: Motion in/on a cone. For motion *on* a cone, the possible orientations of the bond vector are restricted to the surface of the cone. For motion *in* a cone, the vector is orientated anywhere within the cone.

$$S=P_2(\cos\theta_o)=\frac{(3\cos^2\theta_o-1)}{2} \tag{1.65}$$

Here, P_2 is second order Legendre polynomial and θ_o the half-angle of the cone. For wobbling *in* a cone with half opening angle θ_o, S equals (88; 177; 188),

$$S=\tfrac{1}{2}\cos\theta_o[1+\cos\theta_o] \tag{1.66}$$

Wobbling *on* a cone appears an appropriate model for describing the motion of a C-H vector that is caused by rotation around a chemical bond, e.g. rotation of a nucleobase around the glycosidic angle. In fact, restricted diffusion on a cone (half-opening angle θ_o) would be the most appropriate description of such motion, e.g. for motion within the *syn* range; the S^2 equation is then (88):

$$S^2 = [P_2(\cos\theta_o)]^2 + \frac{\sin^2\theta_o \sin^2\gamma}{\gamma^2}[\cos\theta_o + \tfrac{1}{4}\sin^2\theta_o \cos^2\gamma] \qquad (1.67)$$

Here, γ is the half-angular range of the restricted motion. Diffusion *in* a cone is more appropriate when the motion of a C-H vector is caused by a number of rotations about differently orientated chemical bonds.

Alternatively, jump models seem appropriate in many situations, e.g. N- to S-puckering in the ribose ring, gauche/trans transitions of most backbone angles, or *syn/anti* transitions for the glycosidic torsion angle. When a spin vector jumps between different orientations within the molecule during relaxation at an NMR timescale, an N-site jump model can then be chosen (189; 190).

1.6.2 Spin relaxation studies of nucleic acids

Studies of RNA dynamics by NMR are not as widespread as for proteins as pointed out in recent reviews (180; 181; 183; 191-193). Studies of different degrees of comprehensiveness have been carried out on the active-site dynamics of the lead-dependent ribozyme (194-196), the IRES element (197; 198), a tRNA (199), two RNA in free and their protein bound states (182; 200; 201), a group I intron (202), and conformational exchange in an RNA with a non-paired U residue in its helical stem (203). Schwalbe et al. studied in depth the dynamics of the UUCG- (204) and YNMG tetraloops (205). Al-Hashimi et al. introduced the concept of helix elongation to derive domain motion from both NMR spin relaxation and RDC data (120; 179; 206-208) which will be discussed in more detail in the next subsection.

Chowdhury et al. studied the so-called RNA thermometer (209), which has a network of weak hydrogen bonds within a helix and shows resemblance with the apical stem of the encapsidation signal of the hepatitis B virus of the duck that will be discussed in chapters 2 and 4. Destabilization of the RNA structure starts in this region with the onset of heat shock at 42 °C, (209). Chowdhury et al. did not carry out an NMR dynamics study, but presented the structure of this RNA thermometer and investigated its base pair opening by imino exchange. Biophysical data on how decreased RNA stem stability may affect dynamics on different time scales, results from studies of the unstable ε apical stem-loop of the encapsidation signal of the hepatitis B virus of the *duck* in comparison with the stable *human* counterpart (210).

Molecular biology has clearly shown that conformational changes, resulting from such instabilities, play a prominent role in the interaction between proteins and RNA. The importance of conformation changes in interaction between RNA and proteins is underlined by analyses of available RNA-protein structures and their dynamics in free and in bound states. They demonstrate that RNA-target binding almost invariably occurs by mechanisms involving conformational changes rather than rigid lock-and-key docking.

Dynamic binding not only occurs by the well-known mechanism of ‘induced fit’, but also by a second mechanism, namely ‘conformational or tertiary capture’ (179; 206; 207; 211; 212). Secondary structures in RNAs are highly stable, while tertiary structures are less stable. Consequently, helical domains can easily reorient upon target binding (‘induced fit’). This type of binding can be associated with target selectivity that is based on structure, i.e. ‘structural selectivity’ (179; 211), and was argued to dominate when a change in conformation in itself is sufficient to activate a biological process. Alternatively, free RNA has the opportunity to probe a range of different conformations due to the weak tertiary interactions; one or more of the conformations are captured upon binding (‘conformation or tertiary capture’). This mechanism can be associated with ‘dynamic selectivity’ and has been argued to occur when the target is required to be present to activate the biological process. In the classification of RNA-target interaction so far, it is assumed that secondary structures in RNAs are highly stable. In contrast, *duck* ε apical stem-loop is unstable; this instability is essential for binding, implying that a different aspect of RNA dynamics is there involved in the interaction (210).

1.6.3 Residual dipolar couplings and dynamics

Information on dynamics can also be obtained by analyzing RDCs. The idea behind it is that dipole-dipole vectors in motion give rise to dynamically averaged RDC values. In this subsection, we will describe some motional models, and what information on dynamics can be derived given a RDC. A review on this topic is given by Blackledge (213). Furthermore, it will be shown how RDCs can be used to illustrate domain motions.

In section 1.5, expressions for the RDC were derived assuming a fixed orientation of a dipolar vector with respect to the molecular frame, i.e. in the absence of intramolecular dynamics. In the presence of intramolecular dynamics, the expression and derivation of RDCs becomes more complex. If one assumes that intramolecular motions do not affect the alignment tensor, the following expression holds (213; 214):

$$\left\langle D_{IS}\left(\theta,\phi\right)\right\rangle = D_a\left(\left\langle 3\cos^2\left(\theta\right)\right\rangle - 1 + \frac{3}{2}R\left\langle \sin^2\left(\theta\right)\cos\left(2\cdot\phi\right)\right\rangle\right) \tag{1.68}$$

Here θ is the angle between the z-component of the alignment tensor and the bond vector connecting dipoles I and S, and ϕ is the angle between the x-component of the alignment tensor and the bond vector connecting dipoles I and S. The angular brackets denote averaging over conformations relative to the alignment tensor frame. This expression holds for small amplitude local intramolecular dynamics, e.g. small-scale motion of a dipole-dipole vector that preserve the alignment tensor. When the intramolecular motions are of larger amplitude, e.g. helix domain motions, this is likely to change the global shape of the molecule, and thus the alignment tensor is affected (125; 213). The latter situation, which is likely to occur in larger RNAs, will be discussed later in this introduction.

The mean orientation of the vector with respect to the alignment frame, and dynamic averaging with respect to this mean, have a clear influence on the value of the RDC, as in equation (1.68). These intramolecular dynamics are accounted for in the RDC expressions as follows from the derivation below. To express the Hamiltonian for the RDCs in the LAB frame Euler rotations (131) are to be performed, as in section 1.5.1.1. For rigid molecules in the presence of internal dynamics, the Hamiltonian is first transformed from the PAF to local axis frame (LAF) by an Euler rotation: $D\left(\overline{\zeta,\theta,\phi}\right)$. The local axis frame represents a different orientation for each vector, depending on the mean orientation of the vector with respect to the molecular axis frame, and allows the conformational averaging, defined only by the Euler rotations to be considered equivalently for each local frame. Subsequently, the following transformations of the Hamiltonian are carried out: 2) from the LAF to MAF (*D*(α, β, γ), and 3) from the MAF to LAB. Instead of the transformations in equation (1.18), we now have:

$$\hat{H}_{PAF} \xrightarrow{D_{LAF}\left(\overline{\zeta,\theta,\phi}\right)} \hat{H}_{LAF} \xrightarrow{D_{MAF}\left(\alpha,\beta,\gamma\right)} \hat{H}_{MAF} \xrightarrow{D_{LAB}\left(\overline{\alpha'(t),\beta'(t),\gamma'(t)}\right)} \left\langle \hat{H}_{LAB}\right\rangle \tag{1.69}$$

Note that the angles $\left(\overline{\zeta,\theta,\phi}\right)$ represent Euler rotation of dipolar vector from the PAF to the LAF, the overbar denotes a time averaging. The LAF is aligned with the average position around which the intra molecular motions occur. The angles (α, β, γ) represent the Euler rotation from the LAF to the MAF, following the definitions of Brink and Satchler (131). When no dynamics are occurring the first transformation is over Euler angles (0, 0, 0), resulting in a LAF that coincides with the PAF. The equation for RDCS can be rewritten to equation (1.38) for RDCs. The irreducible second rank spherical tensor of the dipolar interaction after the first two rotations expressed in the MAF is:

$$\hat{\hat{R}}\left(\rho_{20}\right)^{MAF} = \sum_{m'=-2}^{+2}\sum_{n'=-2}^{+2} D^{2}_{0m'}\left(\overline{\zeta,\theta,\phi}\right) D^{2}_{n'm'}\left(\alpha,\beta,\gamma\right)\rho_{2m'} \tag{1.70}$$

Here, $\rho_{2m'}$ is the irreducible spherical tensor representation of the dipolar interaction tensor in the LAF (as seen below only the term ρ_{20} is needed). To transform $\rho_{2m'}$ to the LAB frame, a third rotation is required over Euler angles $\left(\overline{\alpha'(t),\beta'(t),\gamma'(t)}\right)$, where the overbar indicates time averaging. The full sets of rotations from PAF to LAB are given by:

$$\hat{\hat{R}}\left(\rho_{2p}\right)^{LAB} = \sum_{m'=-2}^{+2}\sum_{n'=-2}^{+2}\sum_{p'=-2}^{+2} D^{2}_{m'n'}\left(\overline{\zeta,\theta,\phi}\right) D^{2}_{n'n}\left(\alpha,\beta,\gamma\right) D^{2}_{np}\left(\overline{\alpha'(t),\beta'(t),\gamma'(t)}\right)\rho_{2m'} \tag{1.71}$$

Taking into account that only terms $\hat{\hat{R}}\left(\rho_{20}\right)^{LAB}$ and ρ_{20} (in PAF) contribute significantly to the dipolar coupling, equation (1.71) can be written as:

$$\hat{\hat{R}}\left(\rho_{20}\right)^{LAB} = \sum_{n=-2}^{+2}\sum_{n'=-2}^{+2} D^{2}_{on'}\left(\overline{\zeta,\theta,\phi}\right) D^{2}_{n'n}\left(\alpha,\beta,\gamma\right) D^{2}_{no}\left(\overline{\alpha'(t),\beta'(t),\gamma'(t)}\right)\rho_{20} \tag{1.72}$$

Simplifications of this equation can be obtained by imposing specific motional models.

In the presence of local dynamics, where the timescale of local motions is fast compared to overall tumbling of the molecule and these local motions do not affect the overall alignment tensor; a time-average value for the angles (ζ,θ,ϕ) is obtained, i.e. $\left(\overline{\zeta,\theta,\phi}\right) = \left\langle \zeta,\theta,\phi \right\rangle$. The case where local motions are on the same timescale as overall tumbling has partially been worked out by Wong et al. (215). In the following subsections, two specific motion models will be described: diffusion in a cone, and the Gaussian axial fluctuation model.

1.6.3.1 *Diffusion* **in** *a cone and the Gaussian axial fluctuation model*

If we assume a dipole-dipole vector is diffusing *in* a cone (216), the dependence of the RDC on the angle ϕ is completely averaged out, thereby canceling all terms containing ϕ in equation (1.72). After some considerable rewriting, this model of motion significantly simplifies equation (1.72) to (213):

$$\left\langle D(\theta,\phi)\right\rangle = D_a\left\langle\frac{3\cos^2(\theta)-1}{2}\right\rangle\left(\left(\frac{3\cos^2(\beta)-1}{2}\right)+\frac{3}{2}R\cos(2\alpha)\sin^2(\beta)\right) \tag{1.73}$$

Here θ=-$\theta_{average}$, the angle between the z-axis of the local alignment tensor (LAF) to the z-axis of the molecular alignment tensor (MAF). β is the angle between the z-axis of the molecular alignment tensor frame (MAF) and the z-axis of the laboratory frame (LAB), and α is the angle between the x-axes of these frames. In the presence of axially symmetric motion, the above expression can be written as:

$$\left\langle D(\theta,\phi)\right\rangle = S_{axial}D_{IS,static}(\alpha,\beta) \tag{1.74}$$

$$S_{axial}^2 = \left\langle\frac{3\cos^2(\theta)-1}{2}\right\rangle^2 \tag{1.75}$$

Here, S_{axial} is a scaling factor that depends on the amplitude of the motion. θ represents the half-opening angle of the cone in which diffusion is taking place. Using this model, and assuming dynamics do not alter the principal components of the alignment tensor, one can determine the order parameter for each bond vector separately.

In the Gaussian axial fluctuation model (GAF) (217), the angle between the dipole-dipole vector and the z-axis of alignment tensor is assumed to be normally distributed, i.e. a Gaussian distribution, centered on a particular orientation. It is assumed again that motions of the dipole-dipole vector do not affect the alignment tensor, and that motion is fast compared to the overall tumbling of the molecule. An example of motion in which this model would apply, is the libration motion of an aromatic ring in a helix. The aromatic ring is centered on forming a planar basepair with its complementary base in the opposing strand. When considering a N-H bond in such an aromatic ring, it has a specific mean orientation with respect to its alignment tensor described by angle θ_{avg}. The motions that occur, make the ring librate above and below the plane with a certain standard deviation σ. The probability distribution of the bond orientation can be written as:

$$p(\theta)\cdot d\theta = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\theta_{avg})^2}{2\sigma^2}} d\theta \tag{1.76}$$

Equation (1.72) can be expanded using the known trigonometric functions under influence of the GAF-like motions (217).

If we constrain in equation (1.72) such that the average bond vector is orientated along the z-axis of the local alignment tensor (LAF) and libration occurs across the y-axis of the alignment tensor (in that case $\phi=0$ for the entire motion), one obtains after some considerable rewriting the ortho-GAF model (218):

$$\begin{aligned}\langle D_{IS}\rangle = D_a\cdot(s_1\cdot\left(3\cdot\cos^2(\beta)-1\right)+3\cdot s_2\cdot\sin^2(\beta)\cdot\cos(2\cdot\gamma)+\\ \frac{3}{2}\cdot R\cdot\left(s_1\cdot\sin^2(\beta)\cdot\cos(2\cdot\alpha)+2\cdot s_2\cdot\left(\cos^4\left(\frac{\beta}{4}\right)\cdot\cos(2\cdot\delta_1)+\sin^4\left(\frac{\beta}{2}\right)\cdot\cos(2\cdot\delta_2)\right)\right)\end{aligned} \tag{1.77}$$

In which

$$\begin{aligned}\delta_1 &= \alpha+\gamma & \delta_2 &= -\alpha+\gamma\\ s_1 &= 1+3\cdot e^{-2\cdot\sigma^2} & s_2 &= 1-e^{-2\cdot\sigma^2}\end{aligned} \tag{1.78}$$

An analytical expression for the order parameter in the presence of Gaussian fluctuations with amplitude of σ can be derived (217):

$$S^2 = 1-\frac{3}{4}\cdot\left(1-e^{-4\cdot\sigma^2}\right) \tag{1.79}$$

The GAF model is a description of Gaussian motion along one axis. The 3D-GAF is a similar model (219), assuming Gaussian fluctuations over all three of the Cartesian axes.

To obtain the parameters to define the Gaussian fluctuation, an ensemble of structures is calculated. From the ensemble a series of bond orientation with respect to the alignment tensor are obtained, it has an average orientation of θ_{avg} and a standard deviation of σ (220; 221).

1.6.3.2 Visualization of dynamics using residual dipolar couplings

Al-Hashimi et al. (222) developed a method to apply RDCs to visualize spatially correlated motions in HIV-TAR, a RNA structure element consisting of two helices connected via a flexible hinge-like linker.

Studying dynamics of this molecule was initially hampered since helical motions are on the timescale of rotational diffusion. In addition, these helix motions affect the overall shape of the molecule and thus the principal components of the alignment tensor. Elongation of one of the helices made overall tumbling slow when compared to the timescale of helix motions. The long helix provides the alignment axis. In this way, helix motions could be studied by spin relaxation studies (206; 208; 223).

The method of helix elongation has also been applied to HIV-TAR in an alignment medium, allowing studies of dynamics based on RDCs. In addition, the motion of the short helix does not significantly affect the overall shape of the molecule, and therefore the alignment tensor is fixed. This results in an equation for RDCs as in equation (1.73).

Using this equation and the definitions of angles in Figure 1.15A, one may determine the angles α and β, but not γ. α is the twist-angle around the long axis of the first helix, β is the angle between the two helices, and γ is the twist-angle around the long axis of the second helix. To obtain the angle γ the other of the two helices is extended and RDC measurements (207) are performed on the new sample, that measurement allows one to determine the third angle γ. To find structures matching all NMR parameters, a grid-search was performed over the Euler angles. Three conformers could be identified that satisfy all restraints. When connecting the trajectory of the three conformers, the helices rotate around two Euler axes simultaneously; this has been visualized in

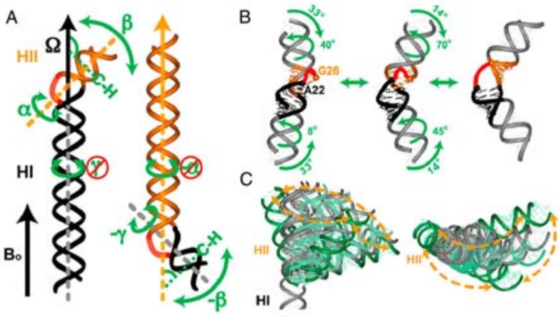


Figure 1.15: Collective helix motions in TAR RNA.

(A) Elongation for visualizing helix motions, of domain and inverse domain, with sensitivity to the rotations twisting (α) and (γ) and bending (β) motions. (B) Transitions between three calculated TAR conformers that give rise to correlated bending and twisting dynamics. (C) Comparison of three TAR ensemble conformers from B (in green) and ligand bound TAR conformations (in gray). Subconformers along the pathway linking the three conformers are shown in light green and the direction of the trajectory is shown in orange arrows. (8)

Figure 1.15.

However, even without helix elongation a possibility does exist to visualize helix motion based on RDCs. In chapter 4 of this thesis, that is demonstrated on a system that is globally similar to HIV-TAR, namely a 62-nucleotide ε element of the hepatitis B virus of the duck. This system consists of two helices with a bulge that acts as a hinge. The structure was calculated and a trajectory of possible alignment tensors and helix orientations was calculated based on RDCs using the method by Wu et al. (125). Refinement of this structure based on RDCs with all possible helix orientations in combination with their alignment tensors resulted in a visual representation of helix motions.

1.7 Appendix I - Rotations in spherical tensor operator space

Generally, the spin Hamiltonian of a given interaction can be written as:

$$\hat{H}_{\lambda} = C^{\lambda}\hat{\mathbf{I}}\cdot\mathbf{A}\cdot\hat{\mathbf{S}} \tag{1.80}$$

Here, $\hat{\mathbf{I}}$ is a Cartesian spin operator, $\hat{\mathbf{S}}$ is the Cartesian vector operator whose exact nature depends on the particular spin interaction, and C^{λ} is an interaction constant (132), which would be for dipolar interaction (130):

$$C^{\lambda} = d = \frac{\mu_0 \hbar \gamma_H \gamma_X}{4\pi r_{HX}{}^3} \tag{1.81}$$

The Hamiltonian in equation (1.80) can be expressed as a sum of scalar products of irreducible spherical tensors of a full rotation point group. There, each scalar product involves an irreducible spin tensor and an irreducible spatial tensor (132).

$$\hat{H}_{\lambda} = C^{\lambda}\sum_{l=0}^{+2}\sum_{m=-l}^{+l}\rho_{lm}^{\lambda}\hat{\mathbf{T}}_{lm}^{\lambda} \tag{1.82}$$

The general definition of the irreducible spherical tensor components ρ_{lm} and $\boldsymbol{T}_{lm}$ can be found in (224). By definition, index *l* is the rank of the tensor, and index *m* the order of the tensor. For the two spins I and S, we have (132):

$$\hat{\mathbf{T}}_{\mathbf{00}} = -\frac{2}{\sqrt{3}}\left(\hat{I}_z\hat{S}_z + \frac{1}{2}\left(\hat{I}_+\hat{S}_- + \hat{I}_-\hat{S}_+\right)\right) \tag{1.83}$$

$$\hat{\mathbf{T}}_{\mathbf{10}} = \frac{1}{\sqrt{2}}\left(\hat{I}_-\hat{S}_+ + \hat{I}_+\hat{S}_-\right) \tag{1.84}$$

$$\hat{\mathbf{T}}_{\mathbf{1\pm1}} = \left(-\hat{I}_{\pm}\hat{S}_z + \hat{I}_z\hat{S}_{\pm}\right) \tag{1.85}$$

$$\hat{\mathbf{T}}_{\mathbf{20}} = \sqrt{\frac{1}{6}}\left(3\hat{I}_z\hat{S}_z - \hat{\mathbf{I}}\cdot\hat{\mathbf{S}}\right) \tag{1.86}$$

$$\hat{\mathbf{T}}_{\mathbf{2\pm1}} = \mp\frac{1}{2}\left(\hat{I}_{\pm}\hat{S}_z + \hat{I}_z\hat{S}_{\pm}\right) \tag{1.87}$$

$$\hat{\mathbf{T}}_{\mathbf{2\pm2}} = \frac{1}{2}\hat{I}_{\pm}\hat{S}_{\pm} \tag{1.88}$$

For the spatial part, the tensor ρ_{lm} is constructed from the irreducible representations of an interaction tensor using the following rules. Consider a rank 2 interaction tensor **A**

$$\mathbf{A} = \begin{pmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{pmatrix} \tag{1.89}$$

This can be decomposed into its irreducible tensors ρ_{lm} as follows (224):

$$\rho_{00} = -\frac{1}{\sqrt{3}}Tr\left(\mathbf{A}\right) = -\frac{1}{\sqrt{3}}\left(A_{xx} + A_{yy} + A_{zz}\right) \tag{1.90}$$

$$\rho_{10} = -\frac{i}{\sqrt{2}}\left(A_{xy} + A_{yx}\right) \tag{1.91}$$

$$\rho_{1\pm1} = -\frac{1}{2}\left(A_{zx} - A_{xz} \pm i\left(A_{zy} - A_{yz}\right)\right) \tag{1.92}$$

$$\rho_{20} = \frac{1}{\sqrt{6}}\left(3A_{zz} - \left(A_{xx} + A_{yy} + A_{zz}\right)\right) \tag{1.93}$$

$$\rho_{2\pm1} = \mp\frac{1}{2}\left(A_{xz} + A_{zx} \pm i\left(A_{yz} - A_{zy}\right)\right) \tag{1.94}$$

$$\rho_{2\pm2} = \frac{1}{2}\left(A_{xx} - A_{yy} \pm i\left(A_{xy} - A_{yx}\right)\right) \tag{1.95}$$

Let us introduce the definitions of the anisotropy parameter δ, and the asymmetry parameter η:

$$\delta = A_{zz} - \frac{1}{3}Tr\left(\mathbf{A}\right) = A_{zz} - \frac{1}{3}\left(A_{xx} + A_{yy} + A_{zz}\right) \tag{1.96}$$

$$\eta = \frac{A_{yy} - A_{xx}}{\delta} = \frac{3\left(A_{xx} - A_{yy}\right)}{A_{xx} + A_{yy} - 2A_{zz}} \tag{1.97}$$

These definitions can be used to calculate the spin Hamiltonian as given in equation (1.82) in any chosen frame. To do so, the high-field approximation is applied, thus the Zeeman Hamiltonian is the dominant term in the full spin Hamiltonian. The only parts of the Hamiltonian are those whose spin parts commute with the operator for the *z* component of the total spin angular momentum I_z (132), i.e. quadrupolar interaction will not be considered here. The only commuting operators in equation (1.82) are $\hat{\mathbf{T}}_{l0}$ and the ranks of interest for NMR are rank 0 and 2, i.e. $l \in \{0,2\}$ (132). Note that terms of rank 0 are invariant under rotations and results in an addition of a constant, equation (1.82) simplifies to:

$$\hat{H} = C^{\lambda}\left(\rho_{00}^{\lambda}\hat{\mathbf{T}}_{00}^{\lambda} + \rho_{20}^{\lambda}\hat{\mathbf{T}}_{20}^{\lambda}\right) \tag{1.98}$$

To express this Hamiltonian in another frame, for example the molecular axis frame, it is convenient to use an Euler rotation over angles (α,β,γ) to transform ρ_{20} to the molecular axis frame. The Euler rotation angles are defined as follows (131); 1) rotate α degrees around the z-axis of the original axis frame; 2) rotate β degrees around the y-axis of the newly obtained axis frame resulting from 1; 3) rotate γ degrees around the z-axis of the axis frame resulting from 2. In spherical coordinates, such a transformation can be written as (133):

$$\hat{\hat{R}}\left(\rho_{lm}\right) = \sum_{m'=-l}^{+l} \rho_{lm'} D_{m'm}^{l}\left(\alpha,\beta,\gamma\right) \tag{1.99}$$

Here $\hat{\hat{R}}\left(\rho_{lm}\right)$ is the superoperator that the irreducible spherical tensor representation ρ_{lm} to a new frame. The Wigner rotation matrices $D_{m'm}^{l}$ used here, are defined as follows:

$$D_{m'm}^{l}\left(\alpha,\beta,\gamma\right) = e^{-i\alpha m'} d_{m'm}^{l}\left(\beta\right) e^{-i\gamma m} \tag{1.100}$$

Here, $d_{m'm}^{l}\left(\beta\right)$ are the coefficients for the reduced Wigner elements (133). In this thesis, the interactions of interest have rank 2, i.e. *l*=2, so we will focus on these interactions.
The reduced Wigner elements of rank 2 ($d_{m'm}^{2}\left(\beta\right)$) read as:

$$d_{m',m}^{2} = \begin{bmatrix} \left(\frac{1+\cos\beta}{2}\right)^2 & -\frac{1+\cos\beta}{2}\sin\beta & \sqrt{\frac{3}{8}}\sin^2\beta & -\frac{1-\cos\beta}{2}\sin\beta & \left(\frac{1-\cos\beta}{2}\right)^2 \\ \frac{1+\cos\beta}{2}\sin\beta & \frac{-1+\cos\beta}{2}+\cos^2\beta & \sqrt{\frac{3}{8}}\sin 2\beta & \frac{1+\cos\beta}{2}-\cos^2\beta & -\frac{1-\cos\beta}{2}\sin\beta \\ \sqrt{\frac{3}{8}}\sin^2\beta & \sqrt{\frac{3}{8}}\sin 2\beta & \frac{3\cos^2\beta-1}{2} & \sqrt{\frac{3}{8}}\sin 2\beta & \sqrt{\frac{3}{8}}\sin^2\beta \\ \frac{1-\cos\beta}{2}\sin\beta & \frac{1+\cos\beta}{2}-\cos^2\beta & \sqrt{\frac{3}{8}}\sin 2\beta & \frac{-1+\cos\beta}{2}+\cos^2\beta & -\frac{1+\cos\beta}{2}\sin\beta \\ \left(\frac{1-\cos\beta}{2}\right)^2 & \frac{1-\cos\beta}{2}\sin\beta & \sqrt{\frac{3}{8}}\sin^2\beta & \frac{1+\cos\beta}{2}\sin\beta & \left(\frac{1+\cos\beta}{2}\right)^2 \end{bmatrix} \tag{1.101}$$

For any additional rotations, equation (1.99) can be further expanded, for example addition of an extra rotation results in:

$$\hat{\hat{R}}\left(\rho_{ln}\right)=\sum_{m'=-l}^{+l}\sum_{n'=-l}^{+l}\rho_{lm'}D^{l}_{m'n'}\left(\alpha,\beta,\gamma\right)D^{l}_{n'n}\left(\alpha',\beta',\gamma'\right) \tag{1.102}$$

Now, if we consider the dipolar interaction, which has no component invariant under rotations (i.e. $\rho_{00}=0$), the spin Hamiltonian in the principal axis frame where the Hamiltonian is diagonalized, equation (1.98) becomes:

$$\hat{H}_{PAF}=d\rho_{20}\hat{\mathbf{T}}_{\mathbf{20}} \tag{1.103}$$

where ρ_{lm} are the irreducible representations of the dipolar coupling tensor **D**, which describes the strength and orientation dependence of the dipole interactions:

$$\mathbf{D}=\begin{pmatrix} D_{xx} & 0 & 0 \\ 0 & D_{yy} & 0 \\ 0 & 0 & D_{zz} \end{pmatrix} \tag{1.104}$$

All terms irreducible representations of **D** are zero except ρ_{20}, because of the axial symmetry of the dipolar coupling tensor **D** and the fact that **D** is traceless.

If we want to transform the Hamiltonian in equation (1.103) to the molecular axis frame (as an example), we need to perform an Euler rotation of ρ_{20} over angles (ζ,θ,ϕ) as described in equation (1.23), so we substitute equation (1.23) and (1.22) with l=2, and m=0:

$$\begin{aligned}\hat{\hat{R}}\left(\rho_{20}\right)=\rho_{2-2}D^{2}_{-20}\left(\zeta,\theta,\phi\right)+\rho_{2-1}D^{2}_{-10}\left(\zeta,\theta,\phi\right)+\rho_{20}D^{2}_{00}\left(\zeta,\theta,\phi\right)+\\ \rho_{21}D^{2}_{10}\left(\zeta,\theta,\phi\right)+\rho_{22}D^{2}_{20}\left(\zeta,\theta,\phi\right)\end{aligned} \tag{1.105}$$

After substitution of D^{2}_{00} using equations (1.100) and (1.101), an ρ_{lm} using equations (1.89) to (1.95), we obtain:

$$\hat{\hat{R}}\left(\rho_{20}\right)=\rho_{20}D^{2}_{00}\left(\zeta,\theta,\phi\right)=\rho_{20}\frac{3\cos^{2}\left(\theta\right)-1}{2} \tag{1.106}$$

So the Hamiltonian in the molecular axis frame evaluates as:

$$\hat{H}_{MAF}=d\,\frac{3\cos^{2}\left(\theta\right)-1}{2}\rho_{20}\hat{\mathbf{T}}_{\mathbf{20}} \tag{1.107}$$

$$\hat{H}_{MAF}=\frac{d}{4}\left(3\cos^{2}\left(\theta\right)-1\right)D_{zz}2\hat{I}_{z}\hat{S}_{z} \tag{1.108}$$

It can be seen that the orientation of the bond vector that defines the z-axis of the dipolar coupling tensor in the principal axis frame, with respect to the z-axis of the molecular axis frame determines (angle θ) the size of the observed dipolar coupling. This provides an example of a transformation of a spin Hamiltonian in a different frame.

1.8 References

1. Al-Hashimi HM, Walter NG. 2008. RNA dynamics: it is about time. *Curr.Op.Struct.Biol.* 18:321-9
2. Ganem D, Prince AM. 2004. Hepatitis B Virus Infection - Natural History and Clinical Consequences. *N Engl J Med* 350:1118-29
3. Altona C, Sundaralingam M. 1972. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J Am Chem Soc* 94:8205-12
4. Watts NR, Conway JF, Cheng N, Stahl SJ, Belnap DM, et al. 2002. The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. *EMBO J.* 21:876-84
5. Beck J, Bartos H, Nassal M. 1997. Experimental Confirmation of a Hepatitis B Virus (HBV) ε-like Bulge-and-Loop Structure in Avian HBV RNA Encapsidation Signals. *Virology* 227:500-4
6. Carter JB, Saunders VA. 2007. Hepadnaviruses (and other reverse-transcribing DNA viruses). In *Virology , Principles and Applications*:213-27. New York: Wiley and Sons. Number of 213-27 pp.
7. Clore GM, Gronenborn AM, Bax A. 1998. A robust method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information. *J Magn Reson* 133:216-21
8. Zhang Q, Al-Hashimi HM. 2009. Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA* 15:1941-8
9. Lu X. 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* 3:1213-27
10. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-21
11. (For a good overview of these Nobel prices, see: http://nobelprize.org)
12. Nowotny M, Yang W. 2009. Structural and functional modules in RNA interference. *Curr.Op.Struct.Biol.* 19:286-93
13. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. *Science* 289:905-20
14. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, et al. 2001. Crystal Structure of the Ribosome at 5.5 A Resolution. *Science* 292:883-96
15. Gibbs WW. 2003. The unseen genome: gems among the junk. *Scientific American* 289:46-53
16. Mandal M, Breaker RR. 2004. Gene Regulation by Riboswitches. *Nature Struct.Biol.* 5:451-63

17. Soukup JK, Soukup GA. 2004. Riboswitches exert genetic control through metabolite-induced conformational change. *Curr.Op.Struct.Biol.* 14 344-9

18. Carter JB, Saunders VA. 2007. *Virology , Principles and Applications*. New York: Wiley and Sons

19. Beck J, Nassal M. 2007. Hepatitis B virus replication. *World J Gastroenterol* 13:48-64

20. Hilleman MR. 2001. Overview of the pathogenesis, prophylaxis and therapeusis of viral hepatitis B, with focus on reduction to practical applications. *Vaccine* 19:1837-48

21. Nachega JB, Chaisson RE. 2003. Tuberculosis Drug Resistance: A Global Threat. *Clinical Infectious Diseases* 36:S24-S30

22. Zhang DK, Chen J, Deng LB, Mao Q, Zheng J, et al. 2010. Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. *Infect Genet Evol* 10:84-8

23. WHO. 2008. *Hepatitis B factsheet 204*. http://www.who.int/mediacentre/factsheets/fs204/en/

24. Hilleman MR. 2003. Critical overview and outlook: pathogenesis, prevention, and treatment of hepatitis and hepatocarcinoma caused by hepatitis B virus. *Vaccine* 21:4626-49

25. Lee WM. 1997. Hepatitis B Virus Infection. *N Engl J Med* 337:1733-45

26. Ellis RW, Kniskern PJ. 1991. Recombinant Hepatitis B vaccines. In *Molecular Biology of the Hepatitis B Virus*, ed. A McLachlan:307-22. Boca Raton, Florida: CRC Press Inc. Number of 307-22 pp.

27. Hilleman MR. 1987. Yeast Recombinant Hepatitus B Vaccine. *Infection* 15:3-7

28. McAleer WJ, Buynak EB, Maigetter RZ, Wampler DE, Miller WJ, Hilleman MR. 1984. Human hepatitis B vaccine from recombinant yeast. *Nature* 307:178-80

29. Monto A, Schooley RT, Lai JC, Sulkowski MS, Chung RT, et al. 2010. Lessons From HIV Therapy Applied to Viral Hepatitis Therapy: Summary of a Workshop. *Am.J.Gastroenterol.* 105:989-1004

30. Damme Pv, Herck Kv. 2007. A review of the long-term protection after hepatitis A and B vaccination. *Travel.Med.Infect.Dis.* 5:79-84

31. Puoti M, Airoldi M, Bruno R, Zanini B, Spinetti A, et al. 2002. Hepatitis B Virus Co-infection in Human Immunodeficiency Virus-infected Subjects. *AIDS Rev.* 4:27-35

32. Schaefer S. 2007. Viral and cellular determinants involved in hepadnaviral entry. *World J Gastroenterol* 13:22-38

33. Nassal M. 2008. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* 134:235-49

34. Ganem D, Varmus HE. 1987. The Molecular Biology of the Hepatitis B Virus. *Ann Rev Biochem* 56 651-93

35. Beck J, Nassal M. 1998. Formation of a Functional Hepatitis B Virus Replication Initiation Complex Involves a Major Structural Alteration in the RNA Template. *Mol.Cell.Biol.* 18 6265-72

36. Beck J, Nassal M. 1997. Sequence- and Structure-Specific Determinants in the Interaction between the RNA Encapsidation Signal and Reverse Transcriptase of Avian Hepatitis B Viruses. *J.Virol.* 71:4971-80

37. Laskus T, Rakela J, Persing DH. 1994. The Stem-Loop Structure of the *cis*-Encapsidation Signal is Highly Conserved in Naturally Occurring Hepatitis B Virus Variants. *Virology* 200:809-12

38. Hu K, Beck J, Nassal M. 2004. SELEX-derived aptamers of the duck hepatitis B virus RNA encapsidation signal distinguish critical and non-critical residues for productive initiation of reverse transcription. *Nucleic Acids Res* 32:4377-89

39. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35:2800-11

40. Beck J, Nassal M. 2001. Reconstitution of a Functional Duck Hepatitis B Virus Replication Initiation Complex from Separate Reverse Transcriptase Domains Expressed in *Escherichia coli*. *J.Virol.* 75:7410-9

41. Hu J, Anselmo D. 2000. In Vitro Reconstitution of a Functional Duck Hepatitis B Virus Reverse Transcriptase: Posttranslational Activation by Hsp90. *J.Virol.* 74:11447–55

42. Lin L, Wan F, Hu J. 2008. Functional and Structural Dynamics of Hepadnavirus Reverse Transcriptase during Protein-Primed Initiation of Reverse Transcription: Effects of Metal Ions. *J.Virol.* 82:5703-14

43. Hu J, Boyer M. 2006. Hepatitis B Virus Reverse Transcriptase and ε RNA Sequences Required for Specific Interaction In Vitro. *J.Virol.* 80:2141-50

44. Hu J, Flores D, Toft D, Wang X, Nguyen D. 2004. Requirement of Heat Shock Protein 90 for Human Hepatitis B Virus Reverse Transcriptase Function. *J.Virol.* 78:13122-31

45. Saenger W. 1988. *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. 556 pp.

46. Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, et al. 1998. Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids (IUPAC Recommendations 1998). *Pure & Appl. Chem.* 70:117-42

47. Stryer L. 1995. Flow of Genetic Information. In *Biochemistry*:95-118. New York: W.H. Freeman and company. Number of 95-118 pp.

48. Varshavsky A. 2006. Discovering the RNA Double Helix and Hybridization. *Cell* 127:1295-7

49. Kolk MH, Graaf Mvd, Wijmenga SS, Pleij CWA, Heus HA, Hilbers CW. 1998. NMR Structure of a Classical Pseudoknot: Interplay of Single- and Double-Stranded RNA. *Science* 280:434-8

50. Liu H, Matsugami A, Katahira M, S. U. 2002. A Dimeric RNA Quadruplex Architecture Comprised of Two G:G(:A):G:G(:A) Hexads, G:G:G:G Tetrads and UUUU Loops. *J Mol Biol* 322:955-70

51. Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499-512

52. Leontis NB, Westhof E. 1998. Conserved geometrical base-pairing patterns in RNA. *Quart. Rev. Biophys.* 31:399-455

53. Tinoco IJ, Bustamante C. 1999. How RNA Folds. *J Mol Biol* 293:273-81

54. Mathews DH. 2006. Predicting RNA secondary structure by free energy minimization. *Theor.Chem.Acc.* 116:160168

55. Banerjee AR, Jaeger JA, Turner CJ. 1993. Thermal Unfolding of a Group I Ribozyme: The Low-Temperature Transition Is Primarily Disruption of Tertiary Structure. *Biochemistry* 32:153-63

56. Liang LG, Draper DE. 1994. Thermodynamics of RNA Folding in a Conserved Ribosomal RNA Domain. *J Mol Biol* 237:560-75
57. Crothers DM, Cole PE, Hilbers CW, Schulman RG. 1974. The Molecular Mechanism of the Thermal Unfolding of Escherichia coli Formylmethionine Transfer RNA. *J Mol Biol* 87:63-72
58. Hilbers CW, Robillard GT, Shulman RG, Blake RD, Webb PK, et al. 1976. Thermal unfolding of yeast glycine transfer RNA. *Biochemistry* 15:1874-82
59. Banerjee AR, Turner DH. 1995. The time dependence of chemical modification reveals slow steps in the folding of a group I ribozyme. *Biochemistry* 34:6504-12
60. Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J Mol Biol* 288:911-40
61. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-15
62. Mathews DH. 2006. Revolutions in RNA Secondary Structure Prediction. *J Mol Biol* 359:526-32
63. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287–92
64. Felden B. 2007. RNA structure: experimental analysis. *Curr.Opin.Microbiol.* 10:286-91
65. Hermann T, Patel DJ. 1999. Stitching Together RNA Tertiary Architectures. *J Mol Biol* 294 829-49
66. Westhof E, Michel F. 1998. Ribozyme architectural diversity made visible. *Science* 282:251-2
67. Ferre-D'Amare AR, Doudna JA. 1999. RNA folds: insights from recent crystal structures. *Annu Rev Biophys Biomol Struct* 28:57-73
68. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, et al. 1981. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc Natl Acad Sci USA* 78:2179-83
69. El Hassan MA, Calladine CR. 1995. The Assessment of the Geometry of Dinucleotide Steps in Double-Helical DNA; a New Local Calculation Scheme. *J Mol Biol* 251:648-64
70. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, et al. 1988. Definitions and nomenclature of nucleic acid structure parameters, "The Cambridge Accord". *EMBO J.* 8:1-4
71. Babcock MS, Pednault EPD, Olson WK. 1994. Nucleic Acid Structure Analysis: Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable Between Structures. *J Mol Biol* 237:125-56
72. Babcock MS, Olson WK. 1994. The Effect of Mathematics and Coordinate System on Comparability and "Dependencies" of Nucleic Acid Structure Parameters. *J Mol Biol* 237:98-124
73. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37
74. Holbrook EL, Holbrook SR. 2001. Crystallization of Nucleic Acids. *Encyclopedia of Life Sciences*:1-4
75. Ke A, Doudna JA. 2004. Crystallization of RNA and RNA–protein complexes. *Methods* 34:408-14
76. Mooers BHM. 2009. Crystallographic studies of DNA and RNA. *Methods* 47:168-76

77. Kovacs H, Moskau D, Spraul M. 2005. Cryogenically cooled probes - a leap in NMR technology. *Prog NMR Spectrosc*:131-55
78. Batey RT, Battiste JL, Williamson JR. 1995. Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol.* 261:300-22
79. Cromsigt JAMTC, Schleucher J, Gustafsson T, Kihlberg J, Wijmenga SS. 2002. Preparation of partially H-2/C-13-labelled RNA for NMR studies. Stereo-specific deuteration of the H5" in nucleotides. *Nucleic Acids Res* 30:1639-45
80. Cromsigt JAMTC, Schleucher J, Kidd-Ljunggren K, Wijmenga SS. 2000. Synthesis of specifically deuterated nucleotides for NMR studies on RNA. *J.Biomol.Struct.Dyn.* Sp.Iss. S2:211-9
81. Nikonowicz EP. 2001. Preparation and use of H-2-labeled RNA oligonucleotides in nuclear magnetic resonance studies. *Methods Enzymol.* 338:320-41
82. Tolbert TJ, Williamson JR. 1996. Preparation of specifically deuterated RNA for NMR studies using a combination of chemical and enzymatic synthesis. *J Am Chem Soc* 118:7929-40
83. Tolbert TJ. 1997. Preparation of specifically deuterated and ^{13}C-labeled RNA for NMR studies using enzymatic synthesis. *J Am Chem Soc* 119:12100-8
84. Nelissen FHT, Gammeren AJv, Tessari M, Girard FC, Heus HA, Wijmenga SS. 2008. Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res* 36:e89
85. Nelissen FHT, Girard FC, Tessari M, Heus HA, Wijmenga SS. 2009. Preparation of selective and segmentally labeled single-stranded DNA for NMR by self-primed PCR and asymmetrical endonuclease double digestion. *Nucleic Acids Res* 37:e114
86. Noggle JS, Schirmer RE. 1971. *The Nuclear Overhauser Effect*. New York: Academic Press
87. Bell RA, Saunders JK. 1970. Correlation of the intramolecular nuclear Overhauser effect with internuclear distance. *Can. J. Chem.* 48:1114-22
88. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387
89. Prestegard JH, Al-Hashimi HM, Tolman JR. 2000. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quart. Rev. Biophys.* 33:371-424
90. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. 1997. Use of Dipolar ^{1}H-^{15}N and ^{1}H-^{13}C Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution. *Nature Struct.Biol.* 4:732-8
91. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. 1995. Nuclear magnetic dipole interactions in field-oriented proteins: Information for structure determination in solution. *Proc Natl Acad Sci USA* 92:9279-83
92. Grishaev A, Ying J, Bax A. 2006. Pseudo-CSA Restraints for NMR Refinements of Nucleic Acid Structure. *J Am Chem Soc* 128:10010-1

93. Ravindranathan S, Kim CH, Bodenhausen G. 2005. Determination of 13 C CSA tensors: Extension of the model-independent approach to an RNA kissing complex undergoing anisotropic rotational diffusion in solution. *J Biomol NMR* 33:163-74

94. Ravindranathan S, Kim CH, Bodenhausen G. 2003. Cross correlations between ^{13}C-^{1}H dipolar interactions and ^{15}N chemical shift anisotropy in nucleic acids. *J Biomol NMR* 27:365-75

95. Fars C, Amata I, Carlomagno T. 2007. C-Detection in RNA Bases: Revealing Structure−Chemical Shift Relationships. *J Am Chem Soc* 129:15814-23

96. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. 2001. Prediction of Proton Chemical Shifts in RNA. Their use in Structure refinement and validation. *J Biomol NMR* 21:11-29

97. Shen Y, Lange O, Delagio F, Rossi P, Aramini JM, et al. 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685-90

98. Shen Y, Vernon R, Baker D, Bax A. 2009. De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* 43:63-78

99. Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G. 2008. CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496-W502

100. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. 2007. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615-20

101. Guntert P. 1998. Structure calculation of biological macromolecules from NMR data. *Quart. Rev. Biophys.* 31:145-237

102. Brunger AT. 1992. *X-PLOR version 3.1 A system for X-ray Crystallography and NMR*. New Haven: Yale University Press

103. Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman M. 1996. Geometric Parameters in Nucleic Acids: Nitrogenous Bases. *J Am Chem Soc* 118:509-18

104. Gelbin A, Schneider B, Clowney L, Hsieh SH, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Am Chem Soc* 118:519-29

105. Scheek RM, Gunsteren WFv, Kaptein R. 1989. Molecular dynamic simulation techniques for determination of molecular structures from NMR data. *Methods Enzymol.* 177:204-18

106. Kirkpatrick S, Gelatt CD, Vecchi JMP. 1983. Optimization by Simulated Annealing. *Science* 220:671-80

107. Nilges M, Clore GM, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* 229:317-24

108. Brunger AT, Nilges M. 1993. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Quart. Rev. Biophys.* 26:49-125

109. Brunger AT, Adams PD, Rice LM. 1997. New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* 17:317-24

110. Hansen MR, Mueller GA, Pardi A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct.Biol.* 5:1065-74

111. Tycko R, Blanco FJ, Ishii Y. 2000. Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole−Dipole Couplings in High-Resolution Biomolecular NMR. *J Am Chem Soc* 122:9340-1

112. Prestegard JH, Bougault CM, Kishore AI. 2004. Residual dipolar couplings in structure determination of biomolecules. *Chem.Rev.* 104:3519-40

113. Gayathri C, Bothner-By AA. 1982. Dipolar Magnetic Field Effects in NMR Spectra of Liquids. *Chem.Phys.Lett.* 87:192-6

114. Buuren BNMv, Schleucher J, Wittmann V, Griesinger C, Schwalbe H, Wijmenga SS. 2004. NMR Spectroscopic Determination of the Solution Structure of a Branched Nucleic Acid from Residual Dipolar Couplings by Using Isotopically Labeled Nucleotides. *Angew Chem Int Ed* 43:187-92

115. Bax A, Tjandra N. 1997. Are Proteins even Floppier than We Thought ? *Nature Struct.Biol.* 4:254-6

116. Prestegard JH. 1998. New techniques in structural NMR anisotropic interactions. *Nature Struct.Biol.* Supplement:517-22

117. Kung HC, Wang KJ, Goljer I, Bolton PH. 1995. Magnetic Alignment of Duplex and Quadruplex DNAs. *J.Magn.Reson.ser.B.* 109:323-5

118. Al-Hashimi HM, Tolman JR, Majumdar A, Gorin A, Patel DJ. 2001. Determining Stoichiometry in Homomultimeric Nucleic Acid Complexes Using Magnetic Field Induced Residual Dipolar Couplings. *J Am Chem Soc* 123:5806-7

119. Al-Hashimi HM, Majumdar A, Gorin A, Kettani A, Skripkin E, Patel DJ. 2001. Field- and Phage-Induced Dipolar Couplings in a Homodimeric DNA Quadruplex: Relative Orientation of G.(C−A) Triad and G-Tetrad Motifs and Direct Determination of C2 Symmetry Axis Orientation. *J Am Chem Soc* 123:633-40

120. Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ. 2002. Concerted Motions in HIV-1 TAR RNA May Allow Access to Bound State Conformations: RNA Dynamics from NMR Residual Dipolar Couplings. *J Mol Biol* 315:95-102

121. Prestegard JH, Tolman JR, Al-Hashimi HM, Andrec M. 1999. Protein Structure and Dynamics from Field-Induced Residual Dipolar Couplings. In *Biological Magnetic Resonance, Volume 17: Structure, Computation and Dynamics in Protein NMR*, ed. NR Krishna, LJ Berliner, 17:311-55. New York: Plenum Publishers. Number of 311-55 pp.

122. Shestakova AK, Makarkina AV, Smirnova OV, Shtern MM, Chertkov VA. 2006. Orientation of molecules by magnetic field as a new source of information on their structures. *Russ.Chem.Bull.* 55:1359-67

123. Plantenga TM, Zijl PCMv, Maclean C. 1982. Studies of quadrupolar and dipolar electric field effects in the NMR spectra of binary mixtures of liquids. *Chem.Phys.* 66:1-9

124. Laws DD, Bitter HML, Jerschow A. 2002. Solid-State NMR Spectroscopic Methods in Chemistry. *Angew Chem Int Ed* 41:3096-129

125. Wu B, Petersen M, Girard FC, Tessari M, Wijmenga SS. 2006. Prediction of molecular alignment of nucleic acids in aligned media. *J Biomol NMR* 35:103-15

126. Clore GM, Gronenborn AM, Tjandra N. 1998. Direct Structure Refinement against Residual Dipolar Couplings in the Presence of Rhombicity of Unknown Magnitude. *J Magn Reson* 131:159-62

127. Moltke S, Grzesiek S. 1999. Structural constraints from residual tensorial coupling in high resolution NMR without and explicit term for the alignment tensor. *J Biomol NMR* 15:77-82

128. Kramer F, Deshmukh MV, Kessler H, Glaser SJ. 2003. Residual Dipolar Coupling Constants: An Elementary Derivation of Key Equations. *Conc Magn Reson* 21A 10-21

129. Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton N. 2007. Averaging of the Spin Hamiltonians and Residual Interactions. In *Protein NMR Spectroscopy; Principles and Practice*:102-13. New York: Academic Press. Number of 102-13 pp.

130. Duer MJ. 2004. Dipolar Coupling: Theory and Uses. In *Introduction to Solid-State NMR Spectroscopy*:151-234. Cambridge: Blackwell Science. Number of 151-234 pp.

131. Brink DM, Satchler GR. 1994. *Angular Momentum*. Oxford: Clarendon Press

132. Duer MJ. 2004. The General Form of a Spin Interaction Hamiltonian in Terms of Spherical Tensors and Spherical Tensor Operators. In *Introduction to Solid-State NMR Spectroscopy*:337-43. Cambridge: Blackwell Science. Number of 337-43 pp.

133. Haeberlen U. 1976. Nuclear Spin Hamiltonian. In *High Resolution NMR in Solids, Selective Averaging*, ed. JS Waugh, Supplement I:5-15. New York: Wiley-VCH. Number of 5-15 pp.

134. Snyder LC. 1965. Analysis of Nuclear Magnetic Resonance Spectra of Molecules in Liquid-Crystal Solvents. *J.Chem.Phys.* 43:4041-50

135. Sandle WJ, Williams OM. 1971. A Classical Approach to Coherence Narrowing during Multiple Scattering of Resonance Radiation (Linear Polarization Case). *J.Phys.B:Atom.Molec.Phys.* 4:531-54

136. Lipsitz RS, Tjandra N. 2004. Residual Dipolar Couplings in NMR Structure Analysis. *Annu Rev Biophys Biomol Struct* 33:387-413

137. Wu Z, Delagio F, Tjandra N, Zhurkin VB, Bax A. 2003. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ^{31}P chemical shift anisotropy. *J Biomol NMR* 26:297-315

138. Bryce DL, Grishaev A, Bax A. 2005. Measurement or Ribose Carbon Chemical Shift Tensor for A-form RNA by liquid crystal NMR spectroscopy. *J Am Chem Soc* 127:7387-96

139. Capelle K. 2006. A Bird’s-Eye View of Density-Functional Theory. *Braz J Phys* 36:1318-46

140. Cinacchi G. 2005. Ordering of biaxial solutes in a smectic solvent. *Chem.Phys.Lett.* 416:238-45

141. Mielke SP, Krishnan VV. 2009. Characterization of protein secondary structure from NMR chemical shifts. *Prog NMR Spectrosc* 54:141-65

142. Grzesiek S, Sass HJ. 2009. From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr.Op.Struct.Biol.* 19:585-95

143. Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* 10:337-50

144. Cornilescu G, Delagio F, Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289-302

145. Dejaegere A, Bryce RA, Case DA. 1999. An empirical analysis of proton chemical shifts in nucleic acids. In *Modeling NMR Chemical Shifts. Gaining Insight into Structure and Environment,* ed. JC Facelli, ACd Dios:194-206. Washington DC: American Chemical Society. Number of 194-206 pp.

146. Wutrich K. 1986. *NMR of Proteins and Nucleic Acids*. New York: John Wiley & Sons. 291 pp.

147. Xu XP, Case DA. 2001. Automated prediction of ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}C' chemical shifts in proteins using a density functional database. *J Biomol NMR* 21:321-33

148. Neal S, Nip AM, Zhang H, Wishart DS. 2003. Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. *J Biomol NMR* 26:215-40

149. Meiler J. 2003. PROSHIFT: Protein chemical shift prediction using artificial neural networks. *J Biomol NMR* 26:25-37

150. Lam SL. 2007. DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acids Res* 35:W713-W7

151. Altona C, Faber DH, Hoekzema AJAW. 2000. Double-helical DNA ^{1}H chemical shifts: an accurate and balanced predictive empirical scheme. *Magn Reson Chem* 38:95-107

152. Grzesiek S, Cordier F, Jaravine V, Barfield M. 2004. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog NMR Spectrosc* 45:275-300

153. Ottink OM, Rampersad SM, Tessari M, Zaman GJR, Heus HA, Wijmenga SS. 2007. Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined–induced fit mechanism. *RNA* 13:2202-12

154. Ying J, Grishaev A, Latham MP, Pardi A, Bax A. 2007. Magnetic field induced residual dipolar couplings of imino groups in nucleic acids from measurements at a single magnetic field. *J Biomol NMR* 39:91-6

155. Ebrahimi M, Rossi P, Rogers C, Harbison GS. 2001. Dependence of ^{13}C NMR Chemical Shifts on Conformations of RNA Nucleosides and Nucleotides. *J Magn Reson* 150:1-9

156. Xu XP, Au-Yeung SCF. 2000. Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods. *J. Phys. Chem. B* 104:5641-50

157. Case DA. 1995. Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* 6:341-6

158. Sitkoff D, Case DA. 1998. Theories of chemical shift anisotropies in proteins and nucleic acids. *Prog NMR Spectrosc* 32 165-90

159. Giessner-Prettre C, Pullman B. 1987. Quantum mechanical calculations of NMR chemical shifts in nucleic acids. *Quart. Rev. Biophys.* 20:113-72

160. De Dios AC, Pearson JG, Oldfield E. 1993. Secondary and Tertiary Structural Effects on Protein NMR Chemical Shifts: An ab Initio Approach. *Science* 260:1491-6

161. Laws DD, de Dios AC, Oldfield E. 1993. NMR chemical shifts and structure refinement in proteins. *J Biomol NMR* 3:607-12

162. Oldfield E. 1995. Chemical shifts and three-dimensional protein structures. *J Biomol NMR* 5:217-25

163. Szilagyi L. 1995. Chemical Shifts in Proteins come of Age. *Prog NMR Spectrosc* 27:325-443

164. Lam SL, Chi LM. 2010. Use of chemical shifts for structural studies of nucleic acids. *Prog NMR Spectrosc* 56:289-310

165. Johnson CE, Bovey FA. 1958. Calculation of Nuclear Magnetic Resonance Spectra of Aromatic Hydrocarbons. *J.Chem.Phys.* 29:1012-4

166. Haigh CW, Mallion RB. 1980. Ring Current Theories in Nuclear Magnetic Resonance. *Prog NMR Spectrosc* 13:303-44

167. Haigh CW, Mallion RB. 1971. Proton magnetic resonance of non-planar condensed benzenoid hydrocarbons I. Spectra of 3,4-benzophenantrene, pentahelicene and hexahelicene. *Mol.Phys.* 22:945-53

168. Haigh CW, Mallion RB. 1971. Proton magnetic resonance of non-planar condensed benzenoid hydrocarbons II. Theory of chemical shifts. *Mol.Phys.* 22:955-70

169. Ribas-Prado R, Giessner-Prettre C. 1981. Parameters for the calculation of the ring current and atomic magnetic anisotropy contributions to magnetic shielding constants: Nucleic acid bases and intercalating agents. *J.Mol.Struct. THEOCHEM* 76:81-92

170. Augspurger JD, Dijkstra CE. 1991. Electromagnetic properties of molecules from a uniform procedure for differentiation of molecular wave functions to high order. *J.Chem.Phys.* 95:9230-8

171. Bertini I, Luchinat C, Parigi G. 2002. Magnetic susceptibility in paramagnetic NMR. *Prog NMR Spectrosc* 40:249-73

172. Bertini I, Luchinat C, Parigi G, Pieratelli R. 2005. NMR Spectroscopy of Paramagnetic Metalloproteins. *ChemBioChem* 6:1536-49

173. Helm L. 2006. Relaxivity in paramagnetic systems: Theory and mechanisms. *Prog NMR Spectrosc* 49:45-64

174. Eichmuller C, Skrynnikov NR. 2007. Observation of μs time-scale protein dynamics in the presence of Ln^{3+} ions: application to the N-terminal domain of cardiac troponin C. *J Biomol NMR* 37:79-95

175. Peng JW, Wagner G. 1992. Mapping of Spectral Density Functions Using Heteronuclear NMR Relaxation Measurements. *J Magn Reson* 98:308-32

176. Fischer MWF, Majumdar A, Zuiderweg ERP. 1998. Protein NMR relaxation: theory, applications and outlook. *Prog NMR Spectrosc* 33:207-72

177. Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY. 2001. NMR studies of Brownian tumbling and internal motions in proteins. *Prog NMR Spectrosc* 38 197-266

178. Wangsness RK, Bloch F. 1953. The dynamical theory of nuclear induction. *Phys.Rev.* 89:728-39

179. Al-Hashimi HM. 2005. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chembiochem* 6:1506-19

180. Latham MP, Brown DJ, McCallum SA, Pardi A. 2005. NMR Methods for Studying the Structure and Dynamics of RNA. *ChemBioChem* 6 1492-505

181. Hall KB. 2008. RNA in Motion. *Curr.Opin.Chem.Biol.* 12:612-8

182. Shajani Z, Drobny G, Varani G. 2007. Binding of U1A protein changes RNA dynamics as observed by C-13 NMR relaxation studies. *Biochemistry* 46:5875-83

183. Al-Hashimi HM. 2007. Beyond static structures of RNA by NMR: Folding, refolding, and dynamics at atomic resolution. *Biopolymers* 86:345-7

184. Werbelow LG, Grant DM. 1977. Intramolecular Dipolar Relaxation in Multispin Systems. *Adv.Magn.Reson.* 9:189-299

185. Lipari G, Szabo A. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J Am Chem Soc* 104:4559-70

186. Lipari G, Szabo A. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J Am Chem Soc* 104:4546-59

187. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. 1990. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J Am Chem Soc* 112:4989-91

188. Kinosita K, Kawato S, Ikegami A. 1997. A theory of fluorescence polarization decay in membranes. *Biophys J* 20:289

189. Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton N. 2007. Relaxation and Dynamic Processes. In *Protein NMR Spectroscopy; Principles and Practice*:333-404. New York: Academic Press. Number of 333-404 pp.

190. Bruschweiler R, Case DA. 1994. Characterisation of Biomolecular Structure and Dynamics by NMR Cross Relaxation. *Prog NMR Spectrosc* 26:27-58

191. Al-Hashimi HM. 2005. RNA functional dynamics and their characterization by solution NMR. *ChemBioChem* 6:1506-19

192. Shajani Z, Varani G. 2007. NMR studies of dynamics in RNA and DNA by C-13 relaxation. *Biopolymers* 86:348-59

193. Getz M, Sun X, Casiano-Negroni A, Zhang Q, Al-Hashimi HM. 2007. NMR Studies of RNA Dynamics and Structural Plasticity Using NMR Residual Dipolar Couplings. *Biopolymers* 86:384-402

194. Lee JH, Pardi A. 2007. Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. *Nucleic Acids Res.* 35:2965-74

195. Hoogstraten CG, Wank JR, Pardi A. 2000. Active site dynamics in the lead-dependent ribozyme. *Biochemistry* 39:9951-8

196. Legault P, Hoogstraten CG, Metlitzky E, Pardi A. 1998. Order, dynamics and metal-binding in the lead-dependent ribozyme. *J. Mol. Biol.* 284:325-35

197. Addess KJ, Basilion JP, Klausner RD, Rouault TA, Pardi A. 1997. Structure and dynamics of the iron responsive element RNA: Implications for binding of the RNA by iron regulatory binding proteins. *J. Mol. Biol.* 274:72-83

198. Showalter SA, Baker NA, Tang CG, Hall K. 2005. Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J. Biomol. NMR* 32:179-93

199. Vermeulen A, McCallum SA, Pardi A. 2005. Comparison of the global structure and dynamics of native and unmodified tRNA. *Biochemistry* 44:6024-33

200. Shajani Z, Varani G. 2005. C-13 NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* 349:699-715

201. Oberstrass FC, Allain FHT, Ravindranathan S. 2008. Changes in dynamics of SRE-RNA on binding to the VTS1p-SAM domain studied by C-13 NMR relaxation. *J. Am. Chem. Soc.* 130:12007-20

202. Eldho NV, Dayie KT. 2007. Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: Implications for catalysis. *J. Mol. Biol.* 365:930-44

203. Blad H, Reiter NJ, Abildgaard F, Markley JL, Butcher SE. 2005. Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J. Mol. Biol.* 353:540-55

204. Duchardt E, Schwalbe H. 2005. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR C-13 relaxation. *J. Biomol. NMR* 32:295-308

205. Ferner J, Villa A, Duchardt E, Widjajakusuma E, Wohnert J, et al. 2008. NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* 36:1928-40

206. Zhang Q, Sun X, Watt ED, Al-Hashimi HM. 2006. Resolving the Motional Modes That Code for RNA Adaption. *Science* 311:653-6

207. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450:1263-7

208. Hansen AL, Al-Hashimi HM. 2007. Dynamics of Large Elongated RNA by NMR Carbon Relaxation. *J Am Chem Soc* 129:16072-82

209. Chowdhury S, Maris C, Allain FHT, Narberhaus F. 2006. Molecular basis for temperature sensing by an RNA thermometer. *Embo J.* 25:2487-97

210. Ampt KAM, Werf RMvd, Nelissen FHT, Tessari M, Wijmenga SS. 2009. The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding. *Biochemistry* 48:10499-508

211. Leulliot N, Varani G. 2001. Current Topics in RNA-Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture. *Biochemistry* 40:7947-56

212. Williamson JR. 2000. Induced fit in RNA–protein recognition. *Nature Struct.Biol.* 7:834-7

213. Blackledge M. 2005. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog NMR Spectrosc* 46:23-61

214. Meiler J, Prompers JJ, Peti W, Griesinger C, Bruschweiler R. 2001. Model-Free Approach to the Dynamic Interpretation of Residual Dipolar Couplings in Globular Proteins. *J Am Chem Soc* 123:6098-107

215. Wong V, Case DA, Szabo A. 2009. Influence of the coupling of interdomain and overall motions on NMR relaxation. *Proc Natl Acad Sci USA* 106:11016–21

216. Woessner DE. 1962. Spin Relaxation Processes in a Two-Proton System Undergoing Anisotropic Reorientation. *J.Chem.Phys.* 36:1-4

217. Bruschweiler R, Wright PE. 1994. NMR Order Parameters of Biomolecules: A New Analytical Representation and Application to the Gaussian Axial Fluctuation Model. *J Am Chem Soc* 116:8426-7

218. Bernado P, Blackledge M. 2004. Anisotropic Small Amplitude Peptide Plane Dynamics in Proteins from Residual Dipolar Couplings. *J Am Chem Soc* 126:4907-20

219. Bouvignies G, Bernado P, Meier S, Cho K, Grzesiek S, et al. 2005. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc Natl Acad Sci USA* 124:13885-90

220. Bouvignies G, Marwick PRL, Blackledge M. 2008. Characterization of protein dynamics from residual dipolar couplings using the three dimensional Gaussian axial fluctuation model. *Proteins* 71:353-63

221. Bouvignies G, Marwick PRL, Bruschweiler R, Blackledge M. 2006. Simultaneous Determination of Protein Backbone Structure and Dynamics from Residual Dipolar Couplings. *J Am Chem Soc* 128:15100-1

222. Chang KY, Tinoco IJ. 1997. The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J Mol Biol* 269:52-66

223. Dethoff EA, Hansen AL, Zhang Q, Al-Hashimi HM. 2010. Variable helix elongation as a tool to modulate RNA alignment and motional couplings. *J Magn Reson* 202:117-21

224. Mehring M. 1983. Irreducible Tensor Representations of Spin Interactions. In *Principles of High Resolution NMR in Solids*:213-8. Berlin: Springer-Verlag. Number of 213-8 pp.

Chapter 2

Unstable part of the Apical Stem of *Duck* Hepatitis B Virus Epsilon shows enhanced Basepair Opening but not Pico- to Nanosecond dynamics and is Essential for Reverse Transcriptase Binding

Ampt, K.A.M.; Werf, R.M. van der; Nelissen, F.H.T.; Tessari, M.; Wijmenga, S. S. (2009):
Biochemistry, **48**(44), pp. 10499-10508

2 Unstable part of the Apical Stem Loop of Duck Hepatitis B Virus Epsilon shows Enhanced Basepair Opening but not Pico- to Nanosecond Dynamics and is Essential for Reverse Transcriptase Binding

2.1 Abstract

Hepatitis B virus (HBV) replication starts with binding of reverse transcriptase (RT) to the apical stem-loop of epsilon, a conserved element of the RNA pre-genome. For *duck* HBV, an *in-vitro* replication system has provided molecular details of this interaction. Further insights can be obtained from the structure and dynamics of the *duck* and *human* apical stem-loops. Previously, we reported these for the *human* apical stem-loop. Here, we present the same for the *duck* counterpart. Unlike its *human* counterpart, the *duck* apical stem is unstable in its middle/upper part and contains non-canonical basepairs. This dynamics study is the first of an unstable RNA/DNA stem. Similar to *human*, the *duck* apical stem comprises two helical segments with a bend angle of ca. 10°, separated by a non-paired mobile U residue. It is capped by a well-structured conserved UGUU loop with two residues mobile on the ps to ns time-scale, one of which is involved in RT binding. Remarkably, the unstable middle/upper part of the stem does not show enhanced ps to ns time scale dynamics. Instead, adenine dispersion relaxation studies indicate enhanced ms time scale dynamics involving basepair opening. It can then be concluded that basepair opening is essential for epsilon-RT binding, because stabilization of the stem abolishes binding. We hypothesize that binding occurs by conformational capture of bases in the basepair open state. The unstable secondary structure of the apical stem-loop makes *duck* epsilon-RT binding unusual in light of recent classifications of RNA target interactions that assume stable secondary structures.

2.2 Introduction

The hepatitis B virus (HBV) is the most common cause of liver infection in the world, infecting mammalian as well as *avian* hosts. Over 300 millions people worldwide are estimated to be chronically infected by HBV and so far no treatment for the efficient elimination of HBV in chronically infected patients exists (1-3). HBV is a member of the *Hepadnaviridae* family consisting of hepatotropic DNA viruses, which also includes related animal viruses such as *duck* HBV (DHBV) and *heron* HBV (HHBV). The hepadnaviral genome replication cycle consists of the following parts: 1) the genome, contained in infectious virions as small, 3.2 kb relaxed circular, partially double-stranded DNA (RC-DNA), enters the infected cell; 2) the RC-DNA is then converted inside the host cell nucleus into a plasmid-like covalently closed circular DNA (cccDNA); 3) genomic and pre-genomic RNAs (pg-RNA) are subsequently transcribed from the cccDNA by cellular RNA polymerase II; 4) viral proteins, like the P-protein (see below) and viral coat proteins are translated from pgRNAs acting as mRNAs; the P-protein consists of a conserved reverse transcriptase domain (RT), a middle spacer, a C-terminal Rnase H domain (RH) and at its N-terminus a terminal domain (TP); 5) the pgRNA contains near its 5'-end an important structurally conserved element, called epsilon (ε); ε is a ca. 60 nucleotide (nt) RNA fragment that folds into a lower stem and apical stem-loop, separated by a ca. 6 nt internal bulge; binding of the RT domain of the P-protein to the apical stem-loop of ε triggers the packaging of the pgRNA into progeny capsids, after which a 4 nt DNA primer is synthesized using the 6 nt internal bulge of ε as a template; subsequently, the P-protein moves to the 3'-end of the pg-RNA, where reverse transcription by the P-protein starts and new RC-DNA is produced (2-5).

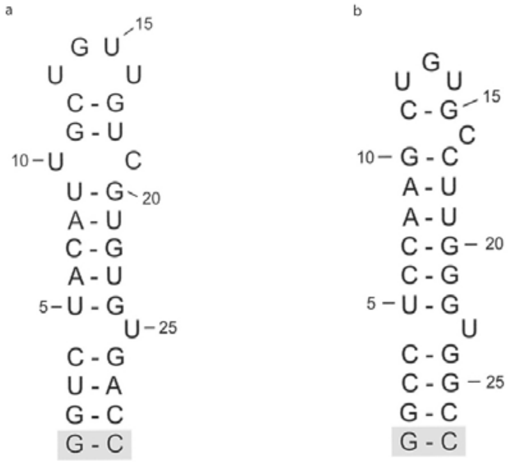


Figure 2.1: Sequence and secondary structure of the apical stem-loop of epsilon of a) duck HBV and b) human HBV. The gray-boxed residues represent structurally silent mutations from the wild type to aid expression

The conserved 60 nt ε RNAs of *human* and *avian* HBV and in particular their apical stem-loops (Figure 2.1) each form a crucial structural element in the replication of their respective HBVs, because binding of the RT domain of the P-protein to the conserved apical stem-loop of ε triggers the molecular events that lead to the reverse transcription of the pre-genomic RNA. Many of the molecular details of the P-ε interaction leading up to the 4 nt primer synthesis have been established, thanks to a fully functional cell-free and chaperone-dependent reconstituted P/ε system based on *duck* HBV (3; 5-10). An *in-vitro* reconstituted system has also been developed for *human* HBV (11; 12). It shows P-ε binding, but no primer synthesis, and provided ε sequence requirements for this binding and demonstrated the involvement of the RT domain in ε binding. Nevertheless, many molecular details of the P-ε interaction stem from the *duck in-vitro* system and a variety of biochemical experiments have been carried out (3; 5-10), including SELEX experiments (9), to define the sequence requirements crucial for binding and/or priming. Based on these biochemical studies, it has been proposed that the replication initiation is a two-step process, where initial physical binding of the ε apical stem-loop to the RT domain is followed by structural rearrangements in ε, leading up to a priming competent complex in which the ε apical stem-loop is opened (9; 13). The *duck* UGUU loop, that caps apical stem, is essential for binding; it contains a U-residue, whose mutation abolishes binding (9). In contrast, the corresponding loop in *human* ε is not to be essential for binding *in-vitro* (12), but is required for encapsidation in intact cells (12). For *heron* HBV, the midsection of the apical stem is not base-paired, while in *duck* HBV, this midsection does contain Watson-Crick base pairs, but is interrupted by non-canonical base pairs. The *duck* (*avian*) *in-vitro* system data now show that full Watson-Crick base pairing (i.e. stabilization) of this part of the *avian* apical stem abolishes or strongly reduces RT-binding (9; 13). In contrast, the *human* ε apical stem is fully Watson-Crick base paired, including its midsection, a feature essential for RT binding (12). A further interesting difference between the two systems concerns the non-paired U in the apical stem (U23 in *human* and U25 in *duck*, Figure 2.1): in *human* ε it is required for RT-binding, but not in *duck (avian)*. On the other hand, in both systems the (right) bottom part of the ε apical stem is needed for binding and appears to interact with the RT. In summary, although the overall secondary structures of *avian* and *human* ε are quite similar as well as the global built of P and also the overall binding and priming processes, in the sequence and structure requirements differences are apparent.

To better understand the molecular details of the crucial RT-ε inactions in both *avian* and *human* HBVs, we set out to determine the structure, dynamics, and thermodynamic properties of their ε RNAs and domains thereof. Previously, we reported the solution structure and dynamics of the *human* ε apical stem-loop (14-16), the thermodynamics of *human* and *avian* (*duck* and *heron)* apical stem-loops and ε-bulges (17), and the secondary - and loop structure of the *duck* apical stem-loop (17). While secondary structure prediction programs predicted a hexaloop capping the apical stem in *human* ε, NMR showed formation of a pseudo triloop (14; 15). The pseudo triloop is an established motif that consists of a hexaloop with trans-loop base pairing between residues 1 and 5 and a bulged out residue 6 (14; 15; 18). The structure of apical stem of *human* ε shows that the non-paired U (U23, Figure 2.1) is partly stacked or resides in the major groove, leading to an angle of ~ 20 ° between the lower and upper part of the apical stem. Here we present the structure of the *duck* apical stem-loop. Dynamic studies on the *human* apical stem-loop showed the capping pseudo triloop to contain highly conserved mobile nucleotides (16). Furthermore, thermodynamic studies (UV melting) have shown the *human* apical stem-loop to be highly stable in contrast to the *avian* apical stem-loops (17). In *human* HBV, the stability of the middle apical stem is essential for ε-RT binding and ultimately priming (see above), whereas in *avian* HBV the instability of the same is essential for RT binding and ultimately priming (see above). The contrasting stabilities of *duck (avian)* and *human* ε apical stems are thus each essential for function. Interestingly, while in most *avian* apical stems, residues in its midsection are not base paired, in wild-type *duck* the apical stem does contain Watson-Crick basepairs, but they are interrupted by non-canonical basepairs, a U:C basepair flanked by two U:G basepairs (Figure 2.1). Our NMR imino melting studies on the wild-type *duck* ε apical stem showed that its midsection, although well formed at lower temperatures, becomes weakened already at 30-40 °C (17). The question then arises how such weakening of the apical stem, which is essential for RT binding (see above), affects its structure and dynamics. We therefore carried out NMR relaxation studies on the *duck* apical stem (-loop) to establish its dynamics on different scales to complement the structural data.

Such a RNA dynamics study is also of biophysical interest, because as far as we are aware, no in-depth dynamics studies have been carried out on an unstable helical RNA A-helical stem containing non-canonical basepairs, such as seen in the *duck* apical stem-loop. RNA dynamics studies by NMR are not as widespread as for proteins as pointed out and evident from recent reviews (19-24). Dynamics studies of different degrees of comprehensiveness have been carried out on the active-site dynamics of the lead-dependent ribozyme (25-27), the IRES element (28; 29), a tRNA (30), two RNA in free and their protein bound states (31-33), a group I intron (34), and conformational exchange in an RNA with a non-paired U residue in its helical stem (35). Schwalbe et al. studied in depth the dynamics of the UUCG- (36) and YNMG tetraloops (37). Al-Hashimi et al. introduced the concept of helix elongation to derive domain motion from either or both NMR spin relaxation or/and RDC data (19; 38; 39). Closest to the *duck* apical stem comes the RNA thermometer studied by Chowdbury et al. (40), which has a network of weak hydrogen bonds within a helix. With the onset of heat shock at 42 °C, destabilization of the RNA structure starts in this region (40). The weakened middle section of the *duck* apical stem also ‘melts’ already at slightly elevated temperatures (30 – 40 °C) (17). Chowdbury et al. (40) presented the structure of this RNA thermometer and investigated its base pair opening by imino exchange, but did not carry out an NMR dynamics study. A dynamics study of the unstable *duck* ε apical stem-loop and comparison with the stable *human* counterpart will thus provide new biophysical data on how decreased RNA stem stability may affect dynamics on different time scales.

Furthermore, a dynamics study of the *duck* apical stem is of interest, because RT binding of the *duck* apical stem involves a weak secondary structure. This is unusual in light of the recent classification of RNA-target interactions as discussed below. The findings from molecular biology clearly show that conformational changes play a prominent role in the interaction between RT and ε leading up to the primer synthesis (9). This is in line with analyses of available RNA-protein structures and their dynamics in free and bound states, which demonstrated that RNA-target binding most commonly occurs by mechanisms involving conformational changes rather than rigid lock-and-key docking; this dynamic binding not only occurs by the well-known mechanism of ‘induced fit’, but also by a second mechanism, namely ‘conformational or tertiary capture’ (19; 38; 39; 41; 42). The reason for the prominence of these dynamic mechanisms is thought to lie in the hierarchical folding of RNA. In contrast to proteins, secondary structures in RNAs are highly stable, whereas tertiary structures are weak. Consequently, helical domains can easily reorient upon target binding (‘induced fit’). This type of binding can be associated with target selectivity that is based on structure, ‘structural selectivity’ (19; 42) and was argued to dominate when a change in conformation in itself is sufficient to activate a biological process. Alternatively, the weak tertiary interactions provide the free RNA with the opportunity to probe a range of different conformations, one or more of which are captured upon binding (‘conformation or tertiary capture’). This mechanism can be associated with ‘dynamic selectivity’ and has been argued to occur when the target is required to be present to activate the biological process. In the above classification of RNA-target interaction, it is assumed that secondary structures in RNAs are highly stable. The *duck* ε apical stem-loop is, in contrast, unstable and this instability is essential for binding, suggesting that a different aspect of RNA dynamics is here involved in the interaction. Dynamics studies on the unstable *duck* ε apical stem-loop and comparison with the stable *human* counterpart will provide new data and broaden the scope of knowledge on dynamic RNA-protein interaction mechanisms.

In this paper, we present the structure and dynamics of the ε apical stem-loop of *duck* HBV determined by NMR spectroscopy. The structure is based on NOEs and dihedral distances complemented with a large number of C-H residual dipolar couplings and is thus of high quality. The dynamics at different time-scales has been determined by means of NMR relaxation. We compare these structure and dynamics data with those from previous studies on the *human* apical stem-loop.

2.3 Materials and Methods

RNA sample preparation. RNA samples were prepared as previously described (17). For RDC measurements two ^{13}C, ^{15}N fully labeled samples in 93:7 H_2O:D_2O with 0.1 mM EDTA, 10 mM Sodium Phosphate buffer, pH 6.7 and concentration 0.72 mM were used. One of the samples was dissolved in 15 mg/ml Pf1 phages (Asla Biotech, Riga, Latvia) for partial alignment to measure C-H residual dipolar couplings (43). The isotropic sample was also used for ^{13}C relaxation measurements.

NMR spectroscopy. NOE and ribose J_{HH}-coupling were measured as described before (17; 44-46) and converted to structure restraints as described (15; 45-47). For RDC measurement HSQC, sometimes in IPAP mode, and 2D and 3D experiments based on HCCH TOCSY were recorded at 15°C to obtain NH and CH couplings (K.A.M. Ampt, R.M. van der Werf; M. Tessari, S.S. Wijmenga, manuscript in preparation). NMR data were processed using NMRPipe and analyzed using Sparky (48; 49). Chemical shift assignments have been submitted to the BMRB database, code 15656 (44).

^{13}C relaxation measurements. ^{13}C T_1, ^{13}C $T_{1\rho}$ and ^{13}C [^{1}H] NOE experiments were measured for C5, C1', C2 and C8 nuclei at 15 °C, on a Varian Unity INOVA600 spectrometer operating at ^{1}H frequency of 599.91 MHz equipped with a triple resonance (^{1}H/^{13}C/^{15}N) probe with XYZ gradients. All experiments were recorded in interleaved fashion as series of 2D (^{13}C, ^{1}H) spectra using sensitivity-enhanced pulse sequences with pulse field gradients (50-53). For C5 and C1' constant time evolution in the indirect dimension was used to prevent peak broadening due to homonuclear ^{13}C coupling. For T_1 and $T_{1\rho}$ experiments 16 scans per FID were acquired using a recovery delay of 2.5 s. For ^{13}C [^{1}H] NOE experiments, 32 scans per FID were acquired, using a recovery delay of 4.9 s. Every relaxation experiment was performed two to five times to check for reproducibility of the results. All T_1 datasets consisted of a series of eight 2D spectra with relaxation times between 20 ms and 800 ms. Similarly, $T_{1\rho}$ datasets consisted of a series of eight 2D spectra with relaxation times between 4 ms and 32 ms. For C2 and C8 a continuous spin lock field of ~ 4.4 kHz was employed. For C1' and C5 a lower spin lock field -1.3 kHz and 2.7 kHz, respectively- was used to minimize Hartmann-Hahn transfer to scalar coupled ^{13}C nuclei (54). Peak intensities were used to fit T_1 and $T_{1\rho}$ relaxation data to exponentially decaying functions. $T_{1\rho}$ values were corrected for resonance offset using standard procedures (36). The influence of ^{13}C-^{13}C interactions on ^{13}C relaxation plays a more significant role with increasing size of RNA molecules. However, as has been shown previously, the C-C interactions in molecules with a similar size as the apical stem-loop studied here, play a minor role (16; 31).

Relaxation dispersion experiments on H2 protons were performed on a Varian Unity INOVA600 spectrometer at 35 ºC and 37.5 ºC (and on a Varian Unity INOVA at 500 MHz at 35 ºC) using a standard relaxation compensated CPMG scheme (55; 56), with a constant-time duration of 15 ms or 20 ms and delay between 180° pulses ranging from 268 μs to 7.5 ms. All experiments were recorded in 1D fashion; for each experiment 1024 scans were acquired with a recovery delay of 3 s.

Structure calculations. Structure calculations were carried out with X-PLOR 3.851 using a torsion angle dynamics protocol (17). Starting from an extended structure, 500 structures were generated using classical NMR restraints (Table S2.1). The twenty lowest energy structures were submitted to a second round of torsion angle dynamics, generating 20 refined structures out of each selected structure. From the resulting 400 refined structures, the 20 with lowest energy were selected for RDC refinement, all with NOE violations < 0.5 Å and with dihedral violations < 5°. For each of these starting structures the rhombicity and the axial component of the alignment tensor were calculated using the method proposed by Wu and co-workers as well as with the software Module (57). Values of Da = -25 Hz and R = 0.08 were used to calculate 5 refined structures from each starting structure resulting in 100 RDC refined structures. Next the stem parts (i.e. residues 1-4, 5-13, 16-24, 26-29) were refined for their too large helical rise (i.e. structures with an average helical rise of 3.5 instead of 2.8 were obtained). This was done by rebuilding the helix with a helical rise of 2.80 and X-displacement of -4.40 in 3DNA (58) and substituting it in the structure, after which the RDC refinement protocol was run again, no restraint violations occurred during this refinement. The final ten lowest energy structures have been submitted to the RCSB protein data bank as 2k5z.

Modelfree analysis. Modelfree 4.20 was used to analyze ^{13}C spin relaxation parameters T_1, T_2 (calculated form $T_{1\rho}$) and heteronuclear ^{13}C NOE. ^{13}C chemical shift anisotropies were taken from Stueber and Grant (C2 = 150 ppm, C5 = 138 ppm and C8 = 134 ppm) and Duchardt and Schwalbe (C1' = 45 ppm) (36; 59). Distances of 1.104 Å for base C-H bonds and 1.115 Å for C1'-H1' bonds were used (54). HydroNMR was used to estimate a value for the rotational correlation time (τ_c) of 9.4 ns (60). This correlation time was used in Modelfree (61) to determine an axial symmetric diffusion model with a τ_c of 8.6 ns and a diffusion anisotropy of 2.7. Model selection was performed as suggested by Mandel et al., using Occam's razor (61). For a few spins the use of the flowchart of Mandel et al. indicated model 3 (determination of S^2 and R_{ex}) should be used. However, the use of model 3 did not seem accurate since no exchange was found when using this model and the flowchart was used to determine a more complicated model.

2.4 Results

In previous studies we have performed structure calculations on the UGUU tetraloop (17). From these calculations it was evident that a U:U basepair is formed in the tetraloop, between residues U13 and U16. This is consistent with the presence of H3 of U13 in the imino region of the NMR spectra. Furthermore, the shift to lower ppm of the H5/6 resonances of U13/16 (44) with respect to reference values (62) shows that they experience ring currents, which indicates stacking consistent with basepair formation. The H3 of U16 could not be detected in the imino spectrum, most likely due to movement of U15 leaving U16 accessible to solvent. In the stem, we find that the H5/6 resonances of U10 and C19 are shifted to lower ppm and thus experience ring current effects. This indicates stacking of these bases and suggests formation of a U10 and C19 non-canonical basepair. In studies made on a segmental labeled sample of full ε, we could observe and identify, thanks to the reduction in overlap, the imino resonance of U10 (U28 in full ε) in the NMR imino spectrum (63). This provides further evidence for the existence of a U10 and C19 basepair. Sugar puckering modes were derived from observed versus not-observed H1’-H2’ cross peaks in the (^{1}H, ^{1}H) DQF COSY spectrum, as well as from C1’H1’ and C2’H2’ couplings (17; 64). All residues are N-puckered, except for the loop residues 14, 15 and 16, which are S-puckered; for residue 25 no pucker could be established.

Structure. Initial structure calculations were carried out with classical restraints, like NOEs and dihedrals, supplemented with hydrogen-bond restraints and weak planarity restraints for base pairs as described below (Table S2.1; see pdb entry 2K5Z for the full list of restraints used in the calculations). Hydrogen-bond restraints were included for canonical base pairs based on experimental evidence, i.e. the presence of imino resonances and their NOE’s. For the two non-canonical base pairs (U13:U16 and U10:C19) hydrogen-bond restraints were included based on the presence of imino resonances and set as loose restraints between U13O2-U16H3, U13H3-U16O4, U10O4-C19H41 and U10H3-C19N3. For the canonical base pairs and for the non-canonical base pair, U10:C19, weak planarity restraints were included. No planarity restraint was put onto the U13:U16 base pair closing the UGUU tetraloop. Starting from an extended structure, a set of 400 structures, of which 115 showed no NOE distance and no dihedral angle violations larger than 0.5 Å and 5°, respectively, was calculated as described in the materials and methods. While NOEs and dihedrals restraints can give valuable information about the local structure it is difficult to obtain information about long distances. It has therefore become increasingly common, to include Residual Dipolar Couplings (RDCs) in structure calculations as they provide information over long distances (20). Aromatic and sugar CH RDCs as well as imino NH RDCs were determined employing phages as alignment medium. Using NOEs and dihedrals restraints complemented with a relatively large number (98) of these RDCs, the 20 lowest energy structures from the initial calculations were refined. RDCs from residues 15 and 16 were not used in the refinement since their narrow and intense resonances suggested that these residues showed local mobility. This was later confirmed by NMR relaxation data (see below). Moreover, their RDC values were small, suggesting averaging out due to internal motion. Indeed when the RDCs of residue 15 and 16 were included in the calculations they consistently showed violations. For instance, when the sugar pucker of residue 16 was left undetermined, the sugar RDCs forced it into N-pucker, even though there is clear evidence, from C1'H1' and C2'H2' J-couplings, that this residue is S-puckered. An overlay of the 10 lowest energy structures is shown in Figure 2.2.

The RMSD of the RDC refined structure set is 0.92 ± 0.43 Å (Table S2.2). The structure is well defined and consists of two helical stems and a UGUU tetraloop (for helical parameters see Table S2.3). The lower stem is separated by the non-paired U, from the upper stem with a bend angle of 10 ± 3 °; the 3 ° standard deviation represents the uncertainty observed in the 10 lowest energy structures and does not represent the amplitude of motion. The non-paired U25, which separates these two stems, is located mostly in the major groove. The upper stem contains a well-structured UC basepair, of which a detailed top view is show in Figure S2.1 (base pair U10:C19). The UGUU tetraloop is well structured with G14 stacked on top of U13. The tetraloop is closed by a U13:U16 base pair, of which a detailed view is shown in

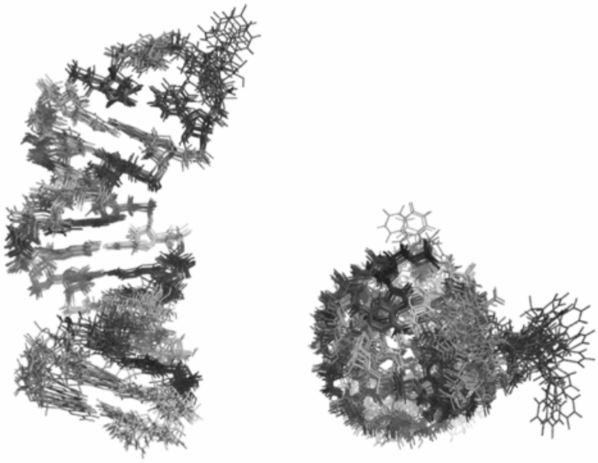


Figure 2.2: *Overlay of the ten lowest energy structures of the duck HBV ε apical stem-loop after RDC refinement (pdb 2k5z).*

Figure S2.1. The loop residue U15, is sticking out of the structure. From the structure calculations a preference for a specific position (straight up, in minor or major groove) of U15 could not be established. Indeed dynamic studies (see below) indicate this residue to be mobile and it, most likely, does not have a specific static position. It is therefore highly solvent accessible in agreement with the absence of the imino resonance in the H_2O imino spectrum (17).

Dynamic studies. The previously described melting studies gave an indication of the dynamics of *duck* apical stem-loop on the ms to s time scale (17). Residual dipolar couplings can give an indication of dynamics on the ps to ms time scale, while relaxation measurements can indicate dynamics on the ps to ns time scale. The relatively small RDCs of residues 15 and 16 already indicated that these residues are mobile. In order to further investigate dynamics on a fast time scale, relaxation studies, T_1, $T_{1\rho}$ and NOE measurements, were carried out for C2, C8, C5 and C1' (Materials and Methods).

In Figure 2.3, the $T_{1\rho}$ relaxation times for C1'-H'1, C2-H2, C5-H5 and C8-H8 are presented. The average $T_{1\rho}$ value for C1'-H1' is 29.6 ms, excluding terminal residues and loop residues. For C2-H2, C5-H5 and C8-H8, the average $T_{1\rho}$ values, again excluding terminal and loop residues, are 23.0 ms, 24.3 ms and 24.7 ms, respectively. Thus, the $T_{1\rho}$ values for the ribose units are on average larger than those for bases, indicating somewhat higher mobility in the ribose units than in the base units. It has to be noted that in the ribose unit, due to overlap, $T_{1\rho}$ values other than those for C1'-H1' were not determined. The loop residues U15 and U16 have ribose $T_{1\rho}$ values that are 15 % larger than the

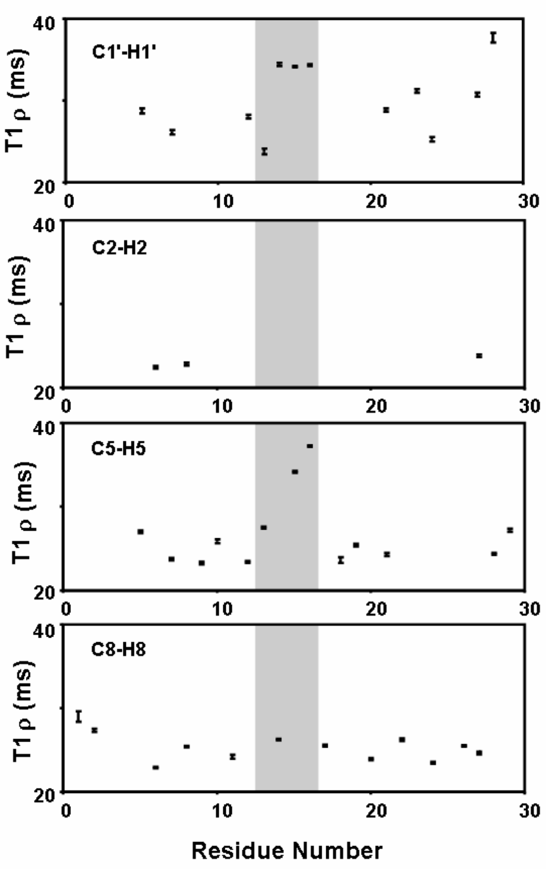


Figure 2.3: *The ^{13}C T1ρ relaxation times for C1', C2, C5, and C8. Resonances that were partially or completely overlapped were excluded from the analysis. Each data point is the average of 2 to 5 separate measurements. The average percentage error is 1 %, based on replicate experiments. The individual errors are indicated as error bars. The gray box indicates the tetraloop.*

average values, while their base $T_{1\rho}$ values are increased by 150 %, indicating that these residues are more flexible on the ps to ns time scale than the other residues. It also suggests that this additional mobility is mostly expressed in their base units. Loop residue G14 has a ribose $T_{1\rho}$ that is 15 % larger than the average, while it’s base $T_{1\rho}$ is similar to the average value. Loop residue U13 has $T_{1\rho}$ values similar to the average values. Shajani and Varani investigated the dynamics of HhaI target DNA and noted that the C1’H1’ resonances of the pyrimidines, especially in the case of deoxycytidines, had a larger transverse relaxation time than the purines (65). Dynamics studies of the DNA dodecamer d(CGCGAATTCGCG)2 also revealed that the C1’-H1’ resonances of the non-terminal deoxycytidines had an increased transverse relaxation time (66). From our RNA relaxation data there is no such distinction observed. In fact, the C1’-H1’ resonances of the cytosine residues have, on average, the same relaxation time as the C1’-H1’ resonances of the other residues. The same observation can be made from relaxation data of the *human* apical stem-loop (16). This suggests that mobility patterns are different in A-helices (RNA) versus B-helices (DNA).

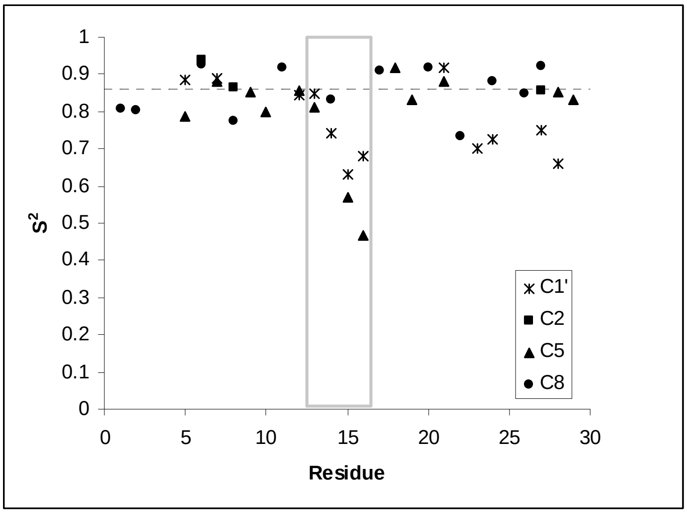


Figure 2.4: *The ^{13}C squared order parameters (see also supplementary table S2.4) as a function of nucleotide sequence. Crosses represent C1’ atoms, squares C2 atoms, triangles C5 atoms, and circles C8 atoms. The dashed line represents the average S^2 value of the stem base C-H groups. A gray box surrounds the loop residues.*

Further analysis was carried out using the program 'Modelfree' (61). The analysis resulted in a correlation time of 8.6 ns and a diffusion anisotropy of 2.7. This agrees reasonably well (within 9 %) with a correlation time of 9.4 ns and a diffusion anisotropy of 2.4 predicted by HydroNMR (60) given the structures derived by RDCs. Full Modelfree results are given in Table S2.4, while Figure 2.4 shows the order parameter, S^2, as a function of nucleotide sequence

The analysis shows that the stem nucleotides are rigid in the base moieties with S^2 values of 0.86 ± 0.06. The ribose moieties in the stems show slightly more motion with S^2 values of

0.80 ± 0.10. These are usual values (16; 36; 54). The time scale of the motions is on average 700 ps. The residues around the non-paired U25, i.e. the opposite residue U5 and the neighboring residue G24, show movement on a slower scale (on average 1000 ps). The loop residues, U13, G14, U15 and U16 can be divided into two sets. U13 and G14 have S^2 values similar to the stem residues, while U15 and U16 show high mobility both for the base and ribose moieties.

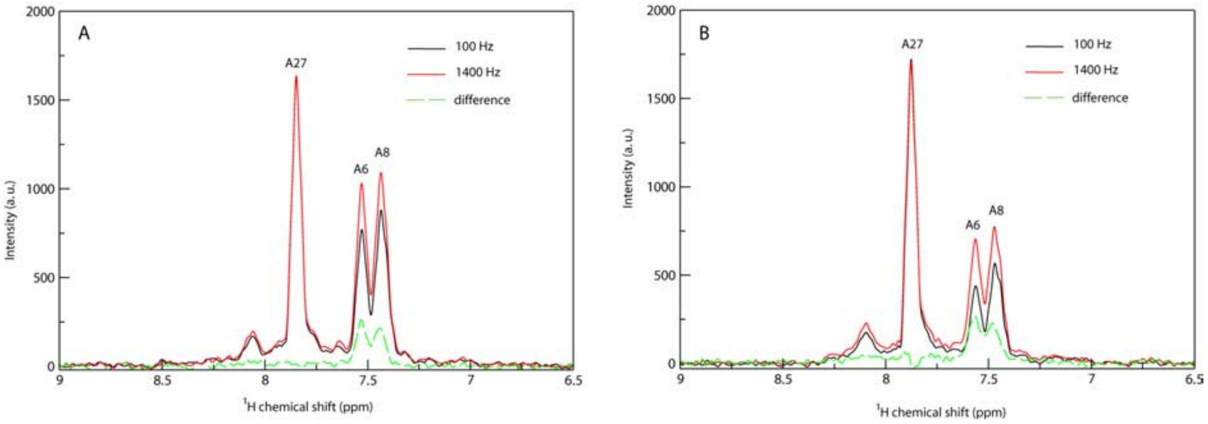


Figure 2.5: *NMR 1D slices from the H2 relaxation dispersion experiments of the 0.72 mM fully $^{13}C/^{15}N$ labeled (10mM Na-Phosphate Ph 6.7, 0.1 mM EDTA, 93:7 H_2O: D_2O) apical stem-loop of duck HBV. The slices are measured at high (1400 Hz, minimal exchange) and low (100 Hz, maximum exchange) effective CPMG field strengths at 35 °C (A) and 37.5 °C (B), respectively.*

To further investigate the dynamics, relaxation dispersion experiments were carried out on H2 spins (see Figure 2.5 and supplementary Figure S2.3). These experiments monitor the presence or absence of conformational exchange on the µs to ms time scale, and, in this case, in the lower stem (A27) and in the middle of the upper stem (A6 and A8). For comparison with previous imino melting studies, the experiments were carried out at 35 and 37.5 °C (17). The difference between the 1D NMR spectra from the dispersion at high (1400 Hz) and low (100 Hz) effective CPMG field (Figure 2.5) shows that residues A6 and A8 in the middle part of the upper stem undergo exchange at these temperatures. No conformational exchange is observed for residue A27 in the lower stem. The full relaxation dispersion curves measured at 500 and 600 MHz at 35 °C and shown in Figure S2.3 for 500 MHz, display the same features, namely exchange for H2 of A6 and A8, but not for H2 of A27. The exchange dispersion curves of H2 of A6/8 are characterized by following fit parameters, 500 MHz : $p_b = 0.024$, $k_{ex} = 506\ s^{-1}$, $R_2^o = 21.5\ s^{-1}$, and $\Delta\omega = 1651$ rad s^{-1} (= 0.53 ppm); 600 MHz : $p_b = 0.045$, $k_{ex} = 645\ s^{-1}$, and $\Delta\omega = 2209$ rad s^{-1} (= 0.59 ppm); the parameters are fully defined in legend of

Figure S2.3. Thus, the H2 of A6 and A8 exchange with a rate (k_{ex}) of 506 - 645 s^{-1} (life time of 1.6 to 2 ms) to a state that is populated to a degree of only 2.4 - 4.5 %. It could further be established that A6 and A8 each exchange with a state in which their H2 resonates at higher chemical shift, i.e. a state in which their H2 experiences less current and thus their bases are less stacked. Thus, A6 and A8 exchange most likely with an unstacked/unfolded state. Note that the $\Delta\omega$ derived from the relaxation dispersion curves has a value of ca. 0.6 ppm, so that in the low-populated (2.4 – 4.5 %) state the H2 of A6/A8 would resonate at ca. 8.1 ppm. In other words, H2 is then in a semi-unstacked state, because in the fully stacked state H2 resonates at 7.5 ppm, while the H2 reference value is 8.4 ppm, which corresponds to a conformation in which it does not experience any ring currents, while an unfolded state would experience some weak ring current effects (circa 8.0 – 8.2 ppm).

2.5 Discussion

The structure of the apical stem-loop of ε *duck* HBV has been derived from NOEs, J-couplings data complemented with long distance information contained in N-H and C-H RDCs. A large number of RDCs (ca. 100) have been used in the final calculations, so that a set of high-quality and well-defined structures is obtained, in which not only local structure features are correct but also long-range features. In addition, this large number of RDCs provides complementary insight into the internal dynamics, which was further assessed based on NMR relaxation data.

The structure set shows the main features of the apical stem-loop. The stem can be divided into two parts, the lower stem, separated from the upper stem by a non-paired U (U25) and is capped by a UGUU tetraloop. The tetraloop is closed by a U:U basepair (U13:U16; see Figure S2.1 for detailed view). The bend angle between the two helical stems was determined to be 10 ± 3°, which is comparable to the bend angle of circa 20°, seen in the structure of the apical stem-loop of *human* HBV. In the calculated structures, the non-paired U25 is located mostly in the major groove. At 25 °C, no resonances could be derived for U25, apart from those of H5 and H6, which were assigned by elimination of all other H5-H6 correlations in the DQF-COSY. The absence of NOE cross-peaks between these resonances and other resonances suggest U25 is not stacked, at least not all the time. In the (^{13}C, ^{1}H) HSQC no H5-C5 or H6-

C6 correlations were detected for U25, suggesting broadening of the C5 (and possibly C6) resonances. Furthermore, at 15 °C these chemical shifts of U25 were not detected in the HSQC and well-resolved 2D and 3D HCN type base-ribose correlation experiments (44). It appears therefore highly likely that all or some of the U25 resonances are broadened beyond detection, due to motional averaging on an intermediate time scale. Finally, we note that in the upper stem a U10:C19 non-canonical basepair is formed (see Figure S2.1 for a detailed view), flanked by two G:U basepairs.

The NMR relaxation data provide a view on the dynamics of the apical stem-loop. As far as we are aware this is the first NMR dynamics study of an RNA molecule with an unstable secondary structure. Most stem residues undergo only slight motion on the ps to ns time scale, i.e. S^2 is 0.86, which corresponds to a C-H fluctuation within a cone with an opening half-angle of circa 20°. The time scale of the fluctuations appear to be sub nanosecond (ca. 700 ps), except for residues neighboring or opposite the non-paired U (U25) and residues neighboring the U10:C19 basepair, which have a somewhat higher degree of fluctuation on a time scale around 1 ns. The S^2 values and time scales are similar to the ones commonly found in stable RNA secondary structures (see e.g. Petzold et al. (16)). In the UGUU tetraloop, the solvent exposed U15 is quite flexible (S^2 ca. 0.6), while U16 in the U13:U16 basepair also shows enhanced motion. In addition, the RDCs of U16 and U15 could be not be fitted in the structure calculation, further confirming their dynamic nature. The other tetraloop residues have S^2 values similar to those of the stem residues. It appears therefore that, apart from these dynamic residues, the apical stem-loop is quite rigid at 15 °C on the ps and ns time scale.

Previously we found that the apical stem-loop of *duck* ε is unstable and partially ‘melts’ already at 35-40 °C as evident form UV melting studies (17). In the apical stem-loop a non-paired U residue (U25) separates the lower and upper stem. The latter contains a number of non-canonical basepairs. NMR temperature studies showed disappearance, already at 35 to 40 °C, of imino resonances from residues in the middle part of the upper stem, but not from residues in the lower stem or directly under the tetraloop, suggesting ‘melting’ of the middle part of the upper stem. It might therefore be hypothesized that such an unstable stem would show increased dynamics. However, we do not observe in our NMR relaxation studies abnormally high internal dynamics in the upper stem on the ps to ns time scale. To further investigate how the instabilities affects dynamics, a temperature profile ranging from 15 °C to ca. 40 °C has been measured of the chemical shifts of the H2 and H8 (Figure S2.2). For H8, no specifically different behavior is seen for the residues in the middle part of the upper stem, apart from G11 next to the U10:C19 basepair and G24/G26 next to the non-paired U25 (as well as G14 in the loop). While H2 of A27 in the lower stem shows no change in chemical shift, a small but significant change in the chemical shift toward higher ppm is observed for H2 of A6 and A8 in the middle part of the upper stem. This indicates an increased presence of non-stacked open states at higher temperatures. Note that conformational shifts of H2 are larger than of H8, so that small structural changes are more readily evident in H2 chemical shifts than H8 chemical shifts. To confirm the presence of conformational exchange on the μs-ms time scale at higher temperatures, relaxation dispersion curves on H2 (C2) have been measured at 35 and 37.5 °C (1D slices of two points on the curve are presented in Figure 2.5 and relaxation dispersion curves are shown in Figure S2.3). We indeed find that residue A27 in the lower stem shows no conformational exchange (see Figure 2.5 and Figure S2.3). On the other hand, residues A6 and A8 in the middle of the upper stem do show conformational exchange to a state in which H2 resonates at higher ppm, consistent with it being an open state. The origin of the instabilities, most apparent in the middle part of the upper stem, lies therefore in enhanced conformational exchange to a semi-unstacked state on the ms time scale, a time scale corresponding to that of base pair opening. Moreover, previous studies have demonstrated that these same residues show increased imino exchange and thus increased base pair opening. It appears therefore appropriate to attribute the enhanced exchange to basepair opening. Surprisingly, no enhanced dynamics on the ps to ns time scales is seen in the unstable upper stem.

How can the structure and dynamics of the free apical stem-loop of *duck* ε provide insight into its binding to the viral RT, an event essential for viral replication? The binding of ε to RT appears to be a two-step process, in which initial physical binding is followed by major conformational changes that lead to a priming competent complex, in which the apical stem-loop of ε has opened up (3; 5; 9; 12). All these events, the initial physical binding and subsequent rearrangements of ε, appear to take place in the palm of the hand of the RT domain. Previously, we reported on the structure (14; 15) and dynamics (16) of the apical stem-loop of the *human* HBV ε. A comparison of the structure and dynamics of the ε apical stem-loops of *duck* HBV, reported here, and *human* HBV can now be made and may provide further insight into the ε -RT interaction. We also discuss how this ε-RT interaction may fit the well-known and recently proposed classes for RNA target interactions.

Comparing the *human* and *duck* ε apical stem-loops several similarities are seen, but also differences stand out. The overall shapes are highly similar. Both have a non-paired U (Figure 2.1: U25, *duck*; U23, *human*), separating the lower and middle/upper parts of the stem. The presence and similar position of the non-paired U leads to a kink in the helix axis in the same direction and of a similar magnitude (bent angles of 10 ± 3 ° and 20 ± 5 ° in *duck* and *human* ε, respectively). The loops capping the ε apical stem-loops are different; in *human* a well-formed pseudo tri-loop is seen, and in *duck* (*avian*) a well-formed UGUU tetraloop. With regard to dynamics we find in both systems relatively high order parameters (S^2 circa 0.9 on average) of the C-H vectors of residues in the lower and upper/middle stem, indicating that the whole kinked apical stem-loop forms in both systems a mainly rigid unit on short time scales. The conservation of the slightly kinked helical shape between *human* and *duck* apical stem-loops, suggests its importance for the interaction with the RT, e.g. it may be required for the apical stem-loop to fit into the palm of the RT. A major difference between the two systems is, that in contrast to *human*, the *duck* apical stem-loop is unstable, mostly in the middle part of the upper stem. This instability enhances conformation exchange involving basepair opening for residues in the middle part of the upper stem as evidenced by our NMR data, but does not affect dynamics on the ps to ns time scales. From our NMR data, we can thus conclude that in *duck* ε open basepair states are present in the middle part of the upper stem, although their population is likely to be low. The ε-RT binding in *avian* HBV requires weak or absent basepairing in the middle of the ε apical stem-loops, i.e. stabilization of this part of the stem abolishes or strongly reduces binding (9). Together this indicates that these open states and thus dynamics form an intrinsic part of the initial binding to RT for the *duck* (*avian*) systems. In view of these characteristics, it is interesting to consider into which of the known classes of RNA target interaction the ε-RT binding in duck and *human* can be categorized.

It has recently been proposed and established that in addition to the well-known class of ‘induced fit’, a second mechanism, ‘conformational or tertiary capture‘, often governs RNA-target interaction(19; 38; 39; 42). The reason for the prominence of these dynamic interaction mechanisms in RNAs is thought to originate from their usually highly stable secondary structures (helices) and weak tertiary structures. Consequently, helical domains can easily reorient upon target binding (‘induced fit’) or probe a range of different conformations in the free RNA, one or more of which are captured upon binding (‘conformation or tertiary capture’). The underlying assumption is thus that RNAs (usually) have stable secondary structures. Indeed, the examples discussed all concern RNAs with stable secondary structures. However, the *duck* ε apical stem-loop has, in contrast, an unstable secondary structure and this instability is essential for RT binding (3; 5; 9; 12) (see also below). This indicates that in the ε-RT interaction a different aspect of RNA dynamics is involved than previously proposed, namely one involving a weak secondary structure instead of or together with a weak tertiary structure. As described and discussed above, from relaxation dispersion and other NMR evidence it can be concluded that the instability leads to enhanced conformational exchange on the ms time scale to an open state, which can be attributed to a basepair open state. Consequently, the *duck* ε-RT binding takes place thanks to enhanced dynamics within an unstable secondary structure element, specifically thanks to enhanced conformational exchange to basepair open states as discussed above, e.g. by capture of bases in the basepair open states.

Note that no base-specific interactions appear to be required for effective *duck* ε-RT binding as follows from, for example, the lines of evidence given below (3; 5; 9; 12). 1) During the SELEX experiments no single winning sequence emerged. 2) From several rounds of SELEX, mutants with a high C-content emerged that still were good binders; the high C-content precludes base pairing. Therefore, a specific sequence is unnecessary for binding in these regions (3; 5; 9; 12). Instead, base pairing that leads to stabilization of the middle part of the upper stem of *duck* ε, abrogates ε-RT binding.

Although basepair open states are essential for binding, it cannot be excluded that the slightly kinked shape of the apical stem-loop aids in the RT interaction. Further, the tetraloop capping the *duck* (or *avian*) ε apical stem, is also essential for RT binding and appears to interact with the RT through a loop residue (U15) since its mutation abrogates binding (9; 13). This residue is highly mobile and could thus be considered to interact through a conformational capture mechanism. In summary, the unstable secondary structure of the ε apical stem-loop makes *duck* ε-RT binding unusual in light of recent classifications of RNA target interactions that assume stable secondary structures. The binding requires exchange to basepair open states and probably occurs by capture of bases in these open states. ‘Tertiary capture’ and ‘Conformational capture’ are used interchangeably in literature, because it is assumed that RNA secondary structures are highly stable. However, in view of our findings, we suggest that a difference should be made between the two. The mechanism for *duck* ε-RT binding, where a weak secondary structure is important for binding, could then be identified as a ‘Secondary Conformational Capture’ mechanism, in contrast to the more common ‘Tertiary Capture’, where a weak tertiary structure is important for binding.

In contrast to *duck*, the apical stem-loop of *human* ε is highly stable (15; 17). As pointed out above, the *human* ε-RT binding is likely to occur by fitting the kinked shape of the apical stem-loop into the palm of the RT. The mechanism could be likened to that of ‘induced-fit’ and/or ‘tertiary-capture’. The non-paired residue U23 (Figure 2.1b) appears to partially insert into the stack of the stem or reside in the major groove, thus inducing small-scale relative motions of the lower versus upper stem (15; 16). Although U23 does show enhanced dynamics on the ps to ns time scales (16), the time scale for partial stacking must be slower than ps to ns and lies probably in the μs to ms range, although no significant exchange broadening was seen for the U23 resonances of the *human* ε apical stem-loop. On the other hand, we do appear to observe in the *duck* ε apical stem-loop exchange broadening for the non-paired U25 (see above).

After initial binding, the apical stem-loop of ε needs to open up (in *human*) or to further open (in *duck*) to form a primer synthesis competent complex (5; 9; 12). The high stability of the *human* ε apical stem-loop represents a significant barrier for opening of the stem. The free energy to overcome this barrier may come from the interaction with the RT, but also from interaction of the capping pseudo triloop with capsid proteins. The pseudo triloop is not needed for binding *in vitro* (12), but is essential for pgRNA encapsidation and initiation of replication in intact cells (5; 12); furthermore, pseudo triloops are known to be often involved in interaction with viral capsid proteins (18).

2.6 Conclusion

The structure determination and the dynamics studies by NMR of the *duck* ε apical stem-loop described here indicate that while its structure is similar to its *human* counterpart, its dynamics are not. The mobility of the structure occurs on different time scales. High mobility on the ps to ns time scales is observed for two residues of the capping UGUU tetraloop comparable to the one highly mobile residue in the capping pseudo triloop in *human*. Remarkably, the unstable middle part of the *duck* ε apical stem-loop displays no enhanced dynamics on the ps to ns time scales, instead it shows enhanced conformational exchange on the ms time scale involving basepair opening. No such conformational exchange is indicated in the stable apical stem-loop of *human* ε. As far as we are aware this is the first NMR dynamics study of an RNA molecule with an unstable secondary structure. Stabilization of the *duck* ε apical stem-loop is known to abolish binding to RT. Dynamics involving the basepair opening in the unstable part of the apical stem-loop is therefore essential for binding to RT. We hypothesize that initial *duck* ε-RT binding involves conformational capture a) of bases in their basepair open states and b) of mobile loop residues. The unstable secondary structure of the ε apical stem-loop makes *duck* ε-RT binding unusual in light of recent classifications of RNA target interactions that assume stable secondary structures. In contrast, initial binding of the highly stable *human* ε to RT likely involves fitting of the stable and kinked apical stem-loop into the palm of the hand of the RT by 'induced fit' or 'conformational or tertiary capture', mechanisms proposed to commonly govern RNA target interactions. In order to confirm our model it would be useful to study the structure of the bound state. Further insight into the ε-RT interaction is also expected to come from the structure and dynamics of the full 60 nt *duck* and *human* ε elements, on which work is to be presented soon.

2.7 References

1. Lee WM. 1997. Medical progress - Hepatitis B virus infection. *New Engl. J. Med.* 337:1733-45
2. Ganem D, Schneider R. 2001. *Fields Virology*. Philadelphia, PA: Lippincott, Williams & Wilkins
3. Nassal M. 2008. Hepatitis B viruses: Reverse transciption a different way. *Virus Research* 134:235-49
4. Nassal M. 2000. *DNA Virus Replication*. Oxford: Oxford University Press. 1-40 pp.

5. Beck J, Nassal M. 2007. Hepatitis B virus replication. *World J. Gastroent.* 13:48-64
6. Beck J, Bartos H, Nassal M. 1997. Experimental confirmation of a hepatitis B virus (HBV) epsilon-like bulge-and-loop structure in avian HBV RNA encapsidation signals. *Virology* 227:500-4
7. Beck J, Nassal M. 2001. Reconstitution of a functional duck hepatitis B virus replication initiation complex from separate reverse transcriptase domains expressed in Escherichia coli. *J. Virol.* 75:7410-9
8. Hu JM, Anselmo D. 2000. In vitro reconstitution of a functional duck hepatitis B virus reverse transcriptase: Posttranslational activation by Hsp90. *J. Virol.* 74:11447-55
9. Hu KH, Beck J, Nassal M. 2004. SELEX-derived aptamers of the duck hepatitis B virus RNA encapsidation signal distinguish critical and non-critical residues for productive initiation of reverse transcription. *Nucleic Acids Res.* 32:4377-89
10. Lin L, Wan F, Hu JM. 2008. Functional and Structural Dynamics of Hepadnavirus Reverse Transcriptase during Protein-Primed Initiation of Reverse Transcription: Effects of Metal Ions *J. Virol.* 82:5703-14
11. Hu JM, Flores D, Toft D, Wang XT, Nguyen D. 2004. Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J. Virol.* 78:13122-31
12. Hu JM, Boyer M. 2006. Hepatitis B virus reverse transcriptase and epsilon RNA sequences required for specific interaction in vitro. *J. Virol.* 80:2141-50
13. Beck J, Nassal M. 1998. Formation of a functional hepatitis B virus replication initiation complex involves a major structural alteration in the RNA template. *Mol. Cell. Biol.* 18:6265-72
14. Flodell S, Schleucher J, Cromsigt J, Ippel H, Kidd-Ljunggren K, Wijmenga S. 2002. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* 30:4803-11
15. Flodell S, Petersen M, Girard F, Zdunek J, Kidd-Ljunggren K, et al. 2006. Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.* 34:4449-57
16. Petzold K, Duchardt E, Flodell S, Larsson G, Kidd-Ljunggren K, et al. 2007. Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res.* 35:6854-61
17. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res.* 35:2800-11
18. Haasnoot PCJ, Brederode FT, Olsthoorn RCL, Bol JF. 2000. A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. *RNA* 6:708-16
19. Al-Hashimi HM. 2005. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chembiochem* 6:1506-19
20. Latham MR, Brown DJ, McCallum SA, Pardi A. 2005. NMR methods for studying the structure and dynamics of RNA. *Chembiochem* 6:1492-505
21. Hall KB. 2008. RNA in motion. *Curr. Opin. Chem. Biol.* 12:612-8

22. Shajani Z, Varani G. 2007. NMR studies of dynamics in RNA and DNA by C-13 relaxation. *Biopolymers* 86:348-59

23. Al-Hashimi HM. 2007. Beyond static structures of RNA by NMR: Folding, refolding, and dynamics at atomic resolution. *Biopolymers* 86:345-7

24. Getz M, Sun XY, Casiano-Negroni A, Zhang Q, Al-Hashimi HM. 2007. NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* 86:384-402

25. Lee JH, Pardi A. 2007. Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. *Nucleic Acids Res.* 35:2965-74

26. Hoogstraten CG, Wank JR, Pardi A. 2000. Active site dynamics in the lead-dependent ribozyme. *Biochemistry* 39:9951-8

27. Legault P, Hoogstraten CG, Metlitzky E, Pardi A. 1998. Order, dynamics and metal-binding in the lead-dependent ribozyme. *J. Mol. Biol.* 284:325-35

28. Addess KJ, Basilion JP, Klausner RD, Rouault TA, Pardi A. 1997. Structure and dynamics of the iron responsive element RNA: Implications for binding of the RNA by iron regulatory binding proteins. *J. Mol. Biol.* 274:72-83

29. Showalter SA, Baker NA, Tang CG, Hall K. 2005. Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J. Biomol. NMR* 32:179-93

30. Vermeulen A, McCallum SA, Pardi A. 2005. Comparison of the global structure and dynamics of native and unmodified tRNA. *Biochemistry* 44:6024-33

31. Shajani Z, Varani G. 2005. C-13 NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* 349:699-715

32. Shajani Z, Drobny G, Varani G. 2007. Binding of U1A protein changes RNA dynamics as observed by C-13 NMR relaxation studies. *Biochemistry* 46:5875-83

33. Oberstrass FC, Allain FHT, Ravindranathan S. 2008. Changes in dynamics of SRE-RNA on binding to the VTS1p-SAM domain studied by C-13 NMR relaxation. *J. Am. Chem. Soc.* 130:12007-20

34. Eldho NV, Dayie KT. 2007. Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: Implications for catalysis. *J. Mol. Biol.* 365:930-44

35. Blad H, Reiter NJ, Abildgaard F, Markley JL, Butcher SE. 2005. Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J. Mol. Biol.* 353:540-55

36. Duchardt E, Schwalbe H. 2005. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR C-13 relaxation. *J. Biomol. NMR* 32:295-308

37. Ferner J, Villa A, Duchardt E, Widjajakusuma E, Wohnert J, et al. 2008. NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* 36:1928-40

38. Zhang Q, Sun XY, Watt ED, Al-Hashimi HM. 2006. Resolving the motional modes that code for RNA adaptation. *Science* 311:653-6

39. Zhang Q, Stelzer, A.C., Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450:1263-7

40. Chowdhury S, Maris C, Allain FHT, Narberhaus F. 2006. Molecular basis for temperature sensing by an RNA thermometer. *Embo J.* 25:2487-97

41. Williamson JR. 2000. Induced fit in RNA-protein recognition. *Nature Struct. Biol.* 7:834-7

42. Leulliot N, Varani G. 2001. Current Topics in RNA-Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture. *Biochemistry* 40:7947-56

43. Hansen MR, Mueller L, Pardi A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* 5:1065-74

44. Ampt KAM, Ottink OM, Girard FC, Nelissen F, Tessari M, Wijmenga SS. 2008. ^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV apical stem loop of the epsilon encapsidation signal. *Biomol. NMR Ass.* 2:159-62

45. Cromsigt J, van Buuren B, Schleucher J, Wijmenga S. 2001. Resonance assignment and structure determination for RNA. *Method Enzymol* 338:371-99

46. Wijmenga SS, van Buuren BNM. 1998. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* 32:287-387

47. Wu B, Girard F, van Buuren B, Schleucher J, Tessari M, Wijmenga S. 2004. Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nucleic Acids Res.* 32:3228-39

48. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* 6:277-93

49. Kneller DG, Kuntz ID. 1993. Ucsf Sparky - an Nmr Display, Annotation and Assignment Tool. *J. Cell. Biochem.*:254-

50. Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, et al. 1994. A General Enhancement Scheme in Heteronuclear Multidimensional Nmr Employing Pulsed-Field Gradients. *J. Biomol. NMR* 4:301-6

51. Yamazaki T, Muhandiram R, Kay LE. 1994. Nmr Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly C-13,N-15-Labeled Proteins - Application to C-13(Alpha) Carbons. *J. Am. Chem. Soc.* 116:8266-78

52. Korzhnev DM, Skrynnikov NR, Millet O, Torchia DA, Kay LE. 2002. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.* 124:10743-53

53. Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY. 2001. NMR studies of Brownian tumbling and internal motions in proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* 38:197-266

54. Hansen AL, Al-Hashimi HM. 2007. Dynamics of large elongated RNA by NMR carbon relaxation. *J. Am. Chem. Soc.* 129:16072-82

55. Loria JP, Rance M, Palmer AG. 1999. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* 121:2331-2

56. Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. 2001. Studying excited states of proteins by NMR spectroscopy. *Nature Struct. Biol.* 8:932-5

57. Wu B, Petersen M, Girard F, Tessari M, Wijmenga SS. 2006. Prediction of molecular alignment of nucleic acids in aligned media. *J. Biomol. NMR* 35:103-15

58. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108-21

59. Stueber D, Grant DM. 2002. C-13 and N-15 chemical shift tensors in adenosine, guanosine dihydrate, 2 '-deoxythymidine, and cytidine. *J. Am. Chem. Soc.* 124:10539-51

60. de la Torre JG, Huertas ML, Carrasco B. 2000. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* 147:138-46

61. Mandel AM, Akke M, Palmer AG. 1995. Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* 246:144-63

62. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. 2001. Prediction of proton chemical shifts in RNA - Their use in structure refinement and validation. *J. Biomol. NMR* 21:11-29

63. Nelissen FHT, van Gammeren AJ, Tessari M, Girard FC, Heus HA, Wijmenga SS. 2008. Multiple segmental and selective isotope labeling of large RNA for NMR structural studies *Nucleic Acids Res.* 36 e89

64. O'Neil-Cabello E, Bryce DL, Nikonowicz EP, Bax A. 2004. Measurement of five dipolar couplings from a single 3D NMR multiplet applied to the study of RNA dynamics. *J. Am. Chem. Soc.* 126:66-7

65. Shajani Z, Varani G. 2008. C-13 relaxation studies of the DNA target sequence for HhaI methyltransferase reveal unique motional properties. *Biochemistry* 47:7617-25

66. Duchardt E, Nilsson L, Schleucher J. 2008. Cytosine ribose flexibility in DNA: a combined NMR C-13 spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res.* 36:4211-9

2.8 Supplementary Material

Table S2.1, restraints used in structure calculations

Structural restraints	Initial calculation	RDC refinement
Distance restraints		
Intraresidue NOE	93	93
Interresidue NOE	139	139
Hydrogen Bonding	59	59
Subtotal	291	291
Torsion Angle restraints		
Glycosidic	28	28
Sugar Pucker	140	140
Backbone Torsion Angles	158	158
Subtotal	326	326
RDC restraints		
CH	-	89
NH	-	9
Subtotal	-	98
Total number of restraints	**617**	**715**

Table S2.2, Structural Statistics

Structural statistics (averaged over 10 lowest energy structures)	RDC refinement
Rms deviation from experimental restraints	
Distance restraints, Å (no violations > 0.5 Å)	0.045 ± 0.004
Dihedral restraints, ° (no violations > 5 °)	0.025 ± 0.004
Residual dipolar couplings, Hz (no violations > 5 Hz)	1.516 ± 0.095
Rms deviation from idealized geometry	
Bonds, Å	0.010 ± 0.000
Angles, °	1.199 ± 0.014
Impropers, °	0.764 ± 0.045
Rms deviation chemical shifts	0.228 ± 0.006
Overall rms deviation (all residues)	
From mean structure, Å	0.917 ± 0.431

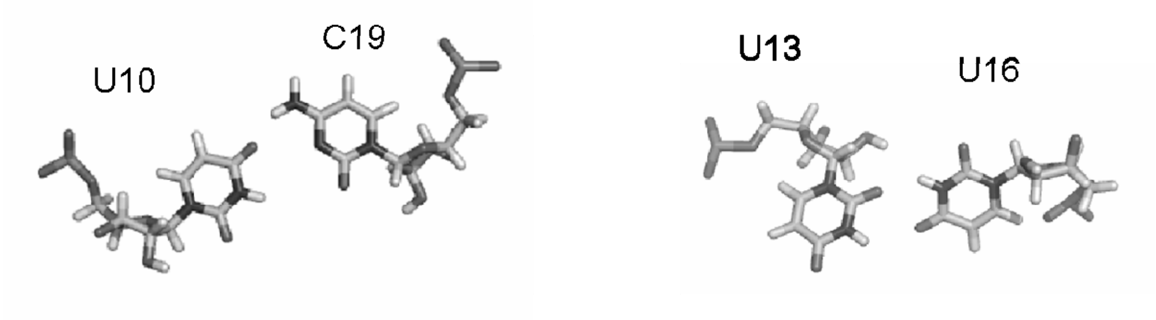


Figure S2.1. Detail top view of the non-canonical base pairs U10-C19 and U13-U16; the most representative examples from structure ensemble were taken.

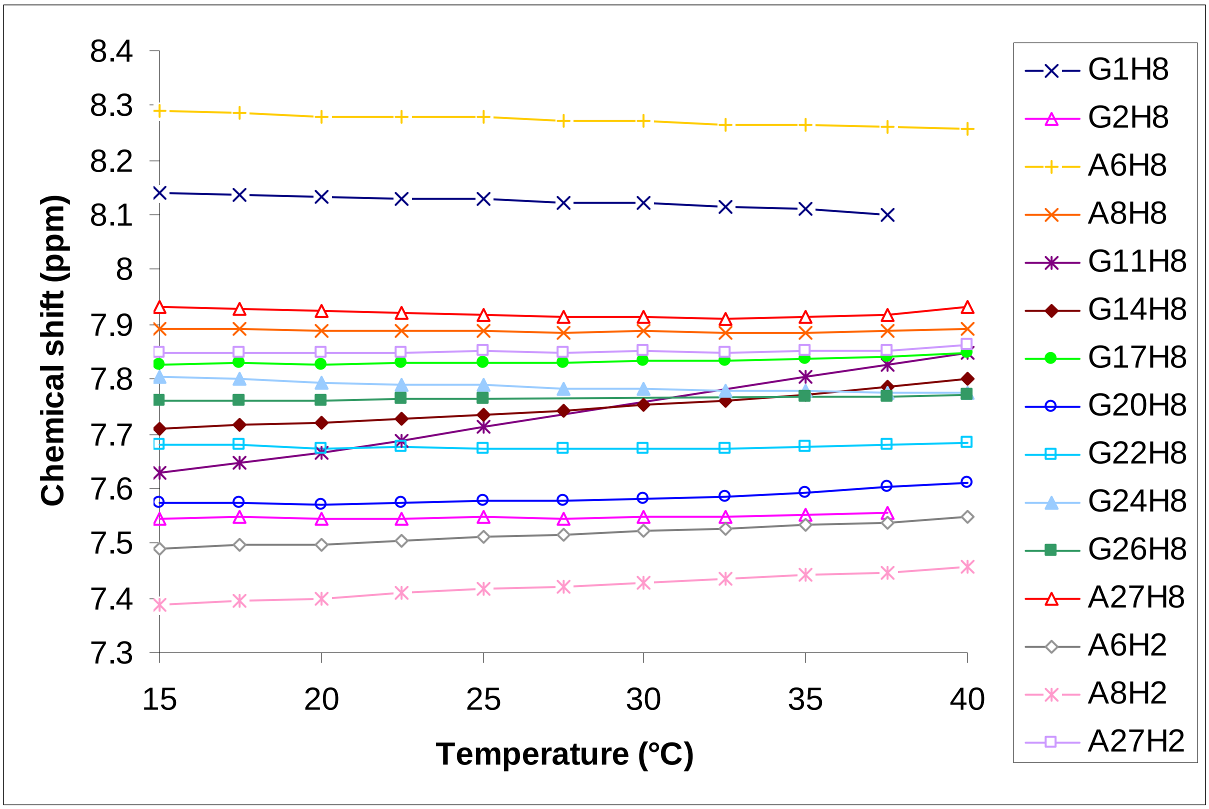


Figure S2.2. Temperature dependence of H2 and H8 chemical shifts at 600 MHz of 0.72 mM fully $^{13}C/^{15}N$-labeled (10mM Na-Phosphate Ph 6.7, 0.1 mM EDTA, 93:7 H_2O: D_2O) apical stem-loop of *Duck* HBV.

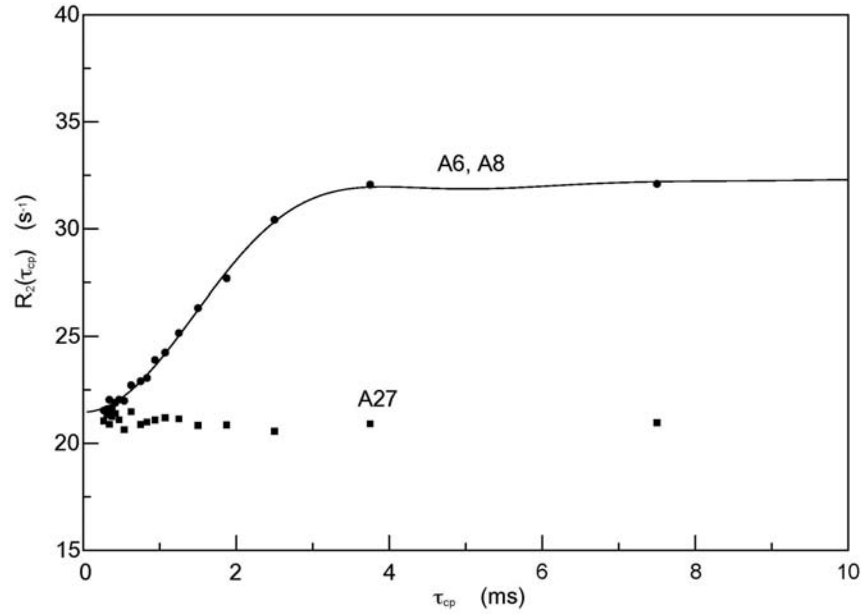


Figure S2.3. Relaxation dispersion plot for the H2 protons of A6 and A8 (filled circles) and A27 (filled squares), recorded with the constant-time relaxation-compensated CPMG sequence (Materials and Methods). The constant-time duration was set to 15 ms and τ_{cp} (the separation between 180° pulses in the CPMG scheme) was increased from 268 μs to 7.5 ms. The experiment was acquired at 35 °C on a Varian Unity INOVA NMR spectrometer operating at 500 MHz proton frequency. The fitted curve results from the simultaneous fitting of R2 (τ_{cp}) for A6 and A8 using the general expression for a two-site exchange by Carver and Richards[$] ($\tau_{cp}$ is the separation between 180° pulses in the pulse CPMG scheme). Fitting parameters: p_b=0.024; k_{ex} = 506 s^{-1}; R_2^o=21.5 s^{-1}; $\Delta\omega$=1651 rad s^{-1}, where, p_b is the population fraction of state b, k_{ex} is the exchange rate in s^{-1}, R_2^o is the exchange-free R_2, and $\Delta\omega$ is the absolute frequency difference between dominant state (state a) and the low populated exchange state (state b). The relaxation dispersion measured at 600 MHz leads similar fit results (see main text, Results section).

[$] J.P. Carver, R.E. Richards, A general two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation, *J. Magn. Reson.* (1972), **6** 89–105.

Table S2.3

Helical Parameters 2k5z

	Shear		Stretch		Stagger		Buckle		Pr-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
G-C	-0.04	0.02	0.06	0.04	-0.19	0.13	-3.13	0.75	-2.16	2.47	-2.47	1.01
G-C	-0.09	0.05	-0.04	0.01	0.58	0.07	1.11	1.27	-12.49	1.06	0.44	0.71
U-A	0.01	0.07	0.05	0.03	0	0.2	6.3	2.63	-6.13	2.78	-1.27	1.48
C-G	0.13	0.1	0.04	0.03	-0.43	0.28	6.17	1.65	-7.4	3.18	0.5	2.71
U-G	2.1	0.04	-0.2	0.07	-0.15	0.34	10.99	2.55	-23.57	4.63	-0.53	2.84
A-U	-0.19	0.19	-0.01	0.06	-0.2	0.13	2.32	2.17	-16.23	2.09	6.63	1.7
C-G	0.06	0.02	0.03	0.04	0.3	0.14	-6.9	0.93	-10.92	1.57	-2.73	1.55
A-U	0.11	0.06	-0.01	0.03	-0.47	0.08	-4.83	1.23	-11.29	1.32	-1.32	1.48
U-G	2.15	0.02	-0.24	0.03	-0.4	0.12	-0.01	0.75	-10.37	2.82	1.87	2.44
U-C	-2.01	0.15	-0.67	0.12	-1.1	0.28	-5.5	4.22	0.09	1.38	-27.86	2.72
G-U	-2.1	0.03	-0.27	0.03	-0.1	0.22	5.71	2.27	-13.43	2.32	-2.47	1.97
C-G	0.14	0.11	-0.23	0.08	-1.13	0.27	8.74	2.24	-10.78	2.41	7.38	1.99
U-U	4.46	1.52	-4.51	2.95	0	1.2	-2.14	15.7	48.18	10.69	52.48	24.52

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
G-C	0	0	0	0	0	0	0	0	0	0	0	0
G-C	-4.4	0	-0.51	0.32	2.8	0	11.01	2.75	6.52	0.81	31.14	1.58
U-A	-4.19	0.12	1.11	0.28	2.75	0.19	15.59	4.31	-5.5	2.28	33.35	1.16
C-G	-2.6	0.95	-1.3	0.73	3.88	0.3	-0.34	6.67	-2.88	2.2	29.09	2.2
U-G	-2.26	0.64	-2.75	0.45	2.79	0.25	12.7	2.33	4.85	3.13	56	5.92
A-U	-4.18	0.31	0.8	0.27	2.9	0.21	25.52	3.65	-4.29	3.17	30.97	1.01
C-G	-4.99	0.2	-0.53	0.23	2.58	0.03	18.82	1.2	6.32	0.71	31.2	1.03
A-U	-3.54	0.25	-0.62	0.3	3.22	0.09	7.67	2.08	-2.42	1.33	31.32	1.11
U-G	-3.78	0.27	0.74	0.38	2.77	0.11	9.27	2	-6.7	1.7	38.29	0.82
U-C	-7.3	0.52	3.51	0.5	1.58	0.19	52.28	3.79	4.71	2.78	29.12	2.98
G-U	-2.67	0.38	-7	1.02	2.72	0.25	1.4	3.52	10.68	4.54	25.63	2.19
C-G	-4.25	0.19	-1.56	0.2	2.76	0.12	13.67	1.68	-1.17	1.16	36.22	0.8
U-U	-0.02	0.44	-3.91	0.99	4.16	0.97	-22.86	6.18	31.44	14.36	68.49	13.62

Table S2.4: Relaxation parameters of ^{13}C nuclei in the apical stem-loop of *Duck* HBV epsilon

Residue	Atom	R_1 (1/s)	R_2 (1/s)	NOE	Model	S^2	S^2_f	S^2_s	τ_e (ps)	R_{ex} (1/s)
1	C8	1.59	36.55	1.253	5	0.81	0.92	0.87	516	
2	C8	1.636	36.55	1.252	5	0.81	0.93	0.87	541	
5	C1'	1.14	34.766		2	0.89		0.89	1053	
	C5	1.908	37.04	1.181	5	0.79	0.96	0.82	1002	
6	C2	1.57	44.494	1.17	2	0.94		0.94	520	
	C8	1.481	43.668	1.18	2	0.93		0.93	680	
7	C1'	1.07	38.228		4	0.89		0.89	936	1
	C5	1.757	42.166	1.23	2	0.88		0.88	665	
8	C2	1.598	43.827	1.158	5	0.87	0.96	0.91	913	
	C8	1.542	39.334	1.178	5	0.78	0.91	0.86	996	
9	C5	1.764	42.941	1.264	2	0.85		0.85	614	
10	C5	2.096	38.589	1.168	2	0.80		0.80	1193	
11	C8	1.566	41.293	1.169	2	0.92		0.92	717	
12	C1'	1.175	35.625		2	0.85		0.85	636	
	C5	1.871	42.708	1.148	2	0.86		0.86	1236	
13	C1'	1.232	42.04		4	0.85		0.85	1565	10
	C5	1.859	36.35	1.329	2	0.81		0.81	390	
14	C1'	1.516	29.049		4	0.74		0.74	307	4
	C8	1.765	38.075	1.354	2	0.83		0.83	381	
15	C1'	1.755	29.317		4	0.63		0.63	317	8
	C5	2.066	29.276	1.323	5	0.57	0.86	0.66	611	
16	C1'	1.772	29.122		4	0.68		0.68	644	4
	C5	2.11	26.831	1.333	5	0.47	0.80	0.59	638	
17	C8	1.587	39.147	1.211	2	0.91		0.91	452	
18	C5	1.645	42.294		2	0.92		0.92	1195	
19	C5	1.92	39.358	1.259	2	0.83		0.83	618	
20	C8	1.547	41.797	1.184	2	0.92		0.92	638	

21	C1'	1.069	34.602		2	0.92		0.92	1080	
	C5	1.74	41.123	1.208	2	0.88		0.88	676	
22	C8	1.514	38.1	1.169	5	0.73	0.87	0.84	1147	
23	C1'	1.111	32.085		2	0.70		0.70	147	
24	C1'	1.414	39.588		4	0.73		0.73	1705	5
	C8	1.625	42.599	1.247	2	0.88		0.88	567	
26	C8	1.524	39.228	1.212	5	0.85	0.94	0.90	606	
27	C1'	1.132	32.572		2	0.75		0.75	179	
	C2	1.619	41.974	1.178	5	0.86	0.95	0.91	708	
	C8	1.542	40.547	1.2	2	0.92		0.92	489	
28	C1'	1.156	26.522		2	0.66		0.66	122	
	C5	1.731	40.999	1.37	2	0.85		0.85	366	
29	C5	1.847	36.763	1.258	2	0.83		0.83	473	

Chapter 3

^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon

Werf, R.M. van der; Girard, F.C.; Nelissen, F.H.T.; Tessari, M.; Wijmenga, S. S. (2009): *Biomol. NMR Assignments*, **2**(2), pp. 143-145

3 ^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon

3.1 Abstract

The replication of the hepatitis B virus is initiated by binding of the viral reverse transciptase protein complex to the apical stem loop of the epsilon element to place it next to the primer loop, from which a 4 nucleotide DNA primer is subsequently synthesized. Here, we present the ^{1}H/^{13}C/^{15}N NMR assignments of the bases and sugars of the 37 residues primer loop of Duck HBV epsilon (BMRB-entry 15786).

3.2 Biological context

The hepatitis B virus is a member of the hepadnavirae (1). It has a small genome of 3.2 kb partially double-stranded DNA, which is replicated after infection via a pregenomic RNA (pg-mRNA) intermediate. The pg-mRNA is encapsidated into immature core particles together with the HBV reverse transcriptase (RT). Binding of the RT protein complex to the apical stem-loop of the encapsidation signal, epsilon (ε), a conserved 60 nt RNA element located at 5’-end of the pregenomic RNA, initiates encapsidation (1; 2). After binding of this complex, a 4 nt DNA primer is synthesized from the internal primer loop of the epsilon signal (1; 3). The resulting complex translocates to a 3’-proximal primer binding site, from where the pg-mRNA is reverse transcribed into the genomic DNA.

More than 350 million people are chronically infected by human HBV (4). No effective treatment exists as yet for chronic infection. Knowledge of the structural details of the HBV replication, specifically of epsilon and RT-epsilon complex, is therefore of crucial importance, for development of anti-virals. Although the replication of human HBV is understood in some detail many questions remain, due to the lack of fully functional in-vitro replication system. Many of the molecular details of HBV replication have instead been worked out from the similar but not identical Duck HBV, for which a fully functional in-vitro replication system does exist (1). We study both the Duck and Human HBV epsilon in parallel and present here the results on the Duck epsilon primer loop.

3.3 Methods and Experiments

The RNA sequence was produced as described earlier by Girard et al. (2). NMR samples of non-labelled RNA with strand concentration of 0.19 mM were prepared in H_2O (10% D_2O) and in D_2O in phosphate buffer (10 mM sodium phosphate pH 6.7, 0.1 mM EDTA); in addition a uniformly $^{13}C/^{15}N$ labelled sample with 1.0 mM RNA strand concentration was prepared in H_2O (10% D_2O) in the same buffer (2).

Sodium-2, 2-Dimethyl-2-silapentane-5-sulfonate (DSS) DSS was added (1.5 mM) to the labelled sample for calibration of ^1H resonances; for unlabeled samples this served as an external reference; ^{13}C and ^{15}N were indirectly calibrated from the ^1H resonances. All experiments were recorded on a Varian Unity INOVA.

Standard methods (5; 6) were used to obtain ^1H resonance assignments and subsequently extended using heteronuclei via (^1H,$^{13}C/^{15}N$) correlation experiments. For this, 500 MHz (^1H, ^1H) NOESY spectra with a 250 ms mixing time were recorded on the non-labelled H_2O sample at 5 °C and at 15 °C. In addition, 800 MHz (^1H,^1H) NOESY spectra with mixing times of 30ms and 500ms were recorded on the non-labelled D_2O sample at 15 °C and 25 °C. Furthermore, a 800 MHz DQF-COSY was recorded at 25 °C (non-labelled D_2O sample) to obtain H5-H6 correlations and sugar puckering from H1'-H2' correlations (2). To obtain ^{13}C resonance assignments, a set of 600 MHz (^{13}C,^1H) HSQC spectra, for the regions C6/H6/C8/H8 (Figure 3.1), C2H2, C5/H5 and C1'/H2', were recorded in H_2O at 15 °C and 25 °C (uniformly labelled sample). To extend the H1'/H2'/C1' sugar resonance assignments to H3'/H4'/C2'/C3'/C4' a 600 MHz a modified 3D HCCH-TOCSY was recorded at 25 °C in H_2O (uniformly labelled sample) (7) and combined with a 600 MHz 2D H1'(C1')C2'-COSY, which correlates H1' and C2' within the same residue, to distinguish C2' from C3' signals. To assign C4 of uridine and C6 of guanine residues from the imino protons a 600 MHz 2D H1/3(N1/3)C4/6 (U/G) experiment was recorded at 15 °C in H_2O (uniformly labelled sample) and subsequently connected to the H5 of uridine via a 600 MHz 2D H5(C5)C4 experiment. The latter experiment also provides H5/C4 correlations in cytosines. Finally, to assign the C2 of uridine and guanine residues a 600 MHz 2D H1/3(N1/3)C2 (U/G) experiment at 15 °C in H_2O (uniformly labelled sample) was employed. The H1/3(N1/3)C2 (U/G) and H1/3(N1/3)C4/6 (U/G) are modified versions of the well known HNCO experiment. To assign N1/3 (U/G) nuclei from imino protons a 600 MHz ^{15}N, ^{1}H-HSQC spectra were recorded at 5 °C and at 15 °C in H_2O (uniformly labelled sample). N1 (U/C) and N9 (G/A) nuclei could be assigned using a 600 MHz 3D H1’C1’N1/9 experiment recorded at 25 °C on the same sample. The heteronuclear experiments mentioned above are (slightly) modified versions of original experiments (see e.g. reviews by Wijmenga and van Buuren and Flinders and Dieckman and references therein) (5; 6)

Spectra were processed using NMRpipe (8) and resonance assignment was performed with Sparky software (9).

3.4 Assignments and data deposition

The assignments at three different temperatures, 5 °C, 15 °C and 25 °C have been deposited into the BMRB databank (accession number 15786). The 250 ms (^{1}H,^{1}H) NOESY in H_2O at 15 °C was recorded to establish the number of base pairs in the RNA molecule and assign the imino proton resonances following sequential walks. All expected imino resonances were observed except for the imino resonances of the potential U7:G31 and A6:U32 at the top of lower stem (Figure 3.1, left); as expected the imino resonances in the internal loop (U7-C13;U30) and for U21 in the UUCG hairpin loop were also not observed (Figure 3.1, left). The H2 resonances were also assigned from the 250 ms (^{1}H,^{1}H) NOESY and later completed/confirmed employing (^{13}C, ^{1}H) HSQCs. Resonances of the other non-exchangeable protons were subsequently assigned via sequential walks (H1’-H6/8/2; H6/8/2-H6/8/2; H5-H6;H2’-H6/8/2) in the 30 ms and 500 ms (^{1}H,^{1}H) NOESY spectra in D_2O combined with (^{1}H,^{1}H) DQF-COSY for the intra-residue H5-H6 correlations. The 500 ms NOESY was used to assign the H1’ protons and H2/H5/H6/H8 protons and the 30 ms NOESY for H2’ and H5/H6/H8 protons. Thanks to the long mixing time (500 ms) H1’-H6/8, H6/8-H6/8 H2’-H6/8) the sequential walks were never completely interrupted. The near complete assignments (see below) of the imino, base and H1’ protons were extended via J-correlated experiments (see Materials and Methods) to the heteronuclei C8/C6/C5 and C2 resonances, N1/3 (G/U), C2(U), C6 (G), N1/9 (U/G). Further assignments of sugar protons and carbons were achieved using the 3D HCCH-TOCSY and 2D H1’(C1’)C2’-COSY; the assignment of the Hx’and Cx’ (x=1’to 4’) were nearly complete in the non-helical regions, while resonances of helix residues were less complete due to the strong overlap (see below); assignment of C5’ and H5’/H5’’was not attempted as these resonances form a CH_2 moiety and were deselected by the delay settings in the 3D HCCH-TOCSY.

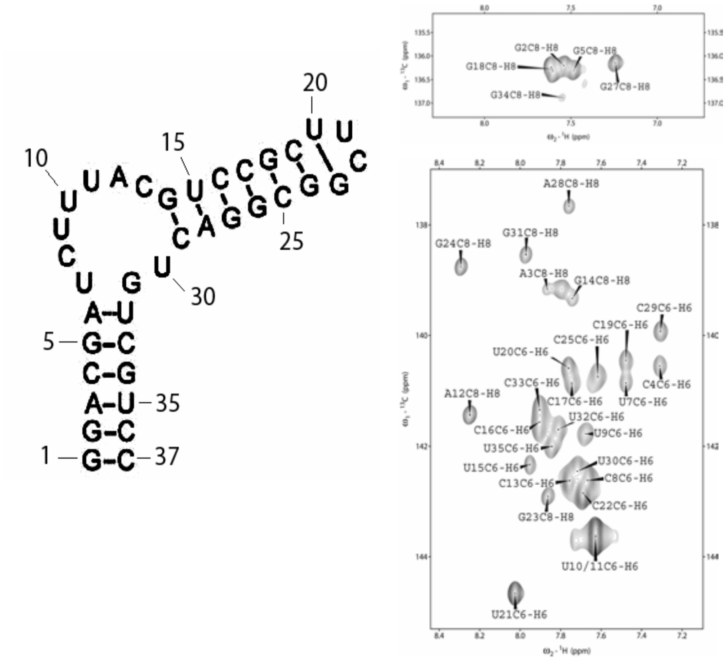


Figure 3.1: 2D $^{1}H/^{13}C$ CT HSQC spectrum recorded at 15°C at 600 MHz on the 1.0 mM sample of the fully $^{13}C/^{15}N$ labelled (10mM sodium phosphate pH 6.7, 0.1 mM EDTA, 1.5 mM DSS, 90:10 H_2O: D_2O) primer loop of Duck HBV. Resonance assignments are indicated by residue type and sequence number followed by atom names. Shown are two regions of the C6H6/C8H8 spectra. On the left the secondary structure of the primer loop is shown. Overlap between U10 and U11 is shown in both the C6 and H6 resonance position

In summary, in the helices all imino proton resonances were assigned as well as their J-correlated N1(G) and N3(U) resonances, only imino proton resonances of basepair A6:U32 and U7:G31 were not observed. A complete assignment of the aromatic protons was obtained, except for the H5 resonance of C36. All C6/8 (A/G) are assigned except for the stem residues G1 C8, A6 C8, G26 C8, C36 C6 and C37 C6 and the internal loop residue A12. All C2(A) are assigned and nine out of ten C4 (U) resonances, nine out of eleven C6 (G) and half the C2(U/G) resonances. The H6 and C6 resonances of U10 and U11 essentially overlap (Figure 3.1, here these resonances overlap, whereas in NOESY spectra the resonances show a slightly different position). The N1(U) and N9 (G) resonances are all assigned, except for N1 for U11, and N9 for G14, G27 and G34. The unambiguous assignments of the sugar resonances (see above), Hx’and Cx’ (x=1’to 4’), are nearly complete in the non-helical regions defined as A6-C13,U20-G24 and U30-U32 (H1’ 100 %;C1’ 69 %; H2’ 100%, C2’ 63 %, H3’ 69 %, C3’ 63 %, H4’ 63 %, C4’ 63 %), while the unambiguous resonance assignment of helix residues is less complete due to the strong overlap, except for H1’ and H2 (ca. 100 %). Assignment of C5’ and H5’/H5’’was not attempted as described above.

Acknowledgements

The 6th framework program of the EU, project FSG-V-RNA, is acknowledged for funding.

3.5 References

1. Beck J, Nassal M. 2007. Hepatitis B virus replication. *World J Gastroenterol* 13:48-64
2. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35:2800-11
3. Flodell S, Petersen M, Girard FC, Zdunek J, Kidd-Ljunggren K, et al. 2006. Solution structure of the apical stem–loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res* 34:4449-57
4. Lee WM. 1997. Hepatitis B Virus Infection. *N Engl J Med* 337:1733-45
5. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387
6. Flinders J, Dieckman T. 2006. NMR spectroscopy of ribonucleic acids. *Prog NMR Spectrosc* 48:137-59
7. Ampt KAM, Wijmenga SS, Tessari M. *Manuscript in preparation*
8. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-93
9. Goddard TD, Kneller DG. 2001. SPARKY 3. San Francisco

Chapter 4

Structural basis for epsilon RNA element-dependent initiation of reverse transcription in hepatitis B viruses

Werf, R.M. van der; Petersen, M.I.P.; Ampt, K.A.M.; Schleucher, J.; Flodell, S.; Girard, F.C.; Nelissen, F.; Wijmenga, S.S.: *Manuscript in preparation*

4 Structural basis for epsilon RNA element-dependent initiation of reverse transcription in hepatitis B viruses

4.1 Abstract

Chronic hepatitis B infection frequently progresses into hepatocellular carcinoma and forms a leading cause of cancer-related mortality worldwide. Current therapies are not effective. Hence there is an urgent need for improved anti-HBV medication, preferably with a different mechanism-of-action. A prerequisite for this is an improved understanding of HBV replication at the molecular level. Hepatitis B replicates its genome via an RNA intermediate. The conserved epsilon element of this pregenomic RNA plays a key-role in this viral replication. Binding of the viral reverse transcriptase to epsilon initiates viral replication by synthesis of the primer using the internal bulge of epsilon as template. We present here the NMR solution structure and dynamics of the 64 nucleotide epsilon elements of human as well as duck hepatitis B virus. The latter was chosen because the duck *in-vitro* replication has provided most of the known molecular details about hepatitis B replication. The NMR solution structures and dynamics unveil similarities and differences between human and duck epsilons and thus indicate differences in the key initial steps in their respective replications mechanisms. However, both epsilons belong to the rare class of asymmetric 6-nt internal bulge loops and display quite similar overall structures. The domain motion of the helices flanking the 6-nt ε primer bulge is remarkably small, like in 5-nt bulges, compared to that of the abundant 3/4-nt bulges. Like the 5-nt bulges of telomerase and HCV, the 6-nt bulge of duck and human HBV ε can be considered structurally encoded functional domains and could form an interesting target for small molecule inhibitors which function by affecting the bulge conformation and/or dynamics. Such inhibitors have already been developed for HCV using the 5-nt bulge as target.

4.2 Introduction

Hepatitis B virus (HBV) is the prototypic member of the hepadnaviridae, a family of small-enveloped DNA viruses infecting mammalian and avian hosts(1). More than 350 million people suffer from chronic HBV infection(2), with chronic hepatitis B frequently progressing into liver cirrhosis and hepatocellular carcinoma, a leading cause of cancer-related mortality worldwide. Despite the availability of an effective prophylactic vaccine(3), the number of chronic carriers has not decreased(4). Current therapies with interferon-alfa or nucleotide analogs generate a sustained response in only about one third of the treated patients, and are accompanied by serious side-effects or rapid development of resistance (5). Hence there is an urgent need for improved anti-HBV medication, preferably with a different mechanism-of-action(6). A prerequisite for this is an improved understanding of HBV replication on the molecular level.

Akin to retroviruses, hepadnaviruses replicate by reverse transcription. However, rather than using a host tRNA as nucleic acid primer, they employ a protein-priming mechanism for initiation (7; 8), in which a unique, additional "Terminal Protein" (TP) domain of the viral reverse transcriptase (P protein) provides a Tyr-residue as extendable end. The highly specific template is an about 60 nt RNA stem-loop, ε, on the pregenomic RNA precursor of the viral DNA that also acts as encapsidation signal (7-10)(Figure 4.1a). Because of severe experimental restrictions with HBV, duck hepatitis B virus (DHBV) has become an invaluable model(11). In particular, DHBV replication initiation has successfully been reconstituted *in vitro*, either in reticulocyte lysate (12) or from recombinant components (13) which, in addition to the reverse transcriptase and ε RNA, must include at least two cellular chaperones, Hsc70 and Hsp40(14; 15). Under appropriate conditions, this leads to copying from the central bulge within ε (primer-bulge / PBD/PBH) of a three or four nt DNA oligo whose 5´ end is covalently attached to P protein via Tyr96 (Tyr63 in HBV) (7; 16-18) (Figure 4.1b). For complete minus-strand DNA synthesis, the oligomer generated in this "priming reaction" is transferred to the 3´ proximal DR1* element on the pregenomic RNA (Figure 4.1c) (8; 19; 20). Specific transfer, essential for proper minus-strand synthesis and thus overall replication, requires that the primer oligomer be copied exactly from the right template sequence in ε; this information must therefore be stored in the specific 3D structure of the P protein - ε RNA complex.

Due to its virtual insolubility and chaperone-dependence, no direct structural information is available for the P protein, except for homology-modeling based predictions (21). For ε RNA, biochemical and mutational analyses (22) have provided some information on the interaction with P protein. Accordingly, formation of a priming-active complex appears to be a two-step process in which an initial binding event is followed by a rearrangement in the RNA (and in the protein) during which the upper stem melts and the RNA gains a template-active conformation (23), supposedly by an induced-fit mechanism. Critical features beyond the PBD (which *must* be brought into P protein’s active site) are the apical loop, the two- to three base-pairs of the lower stem immediately underneath the PBD (24), and likely some residues in the upper stem (9). For HBV, specific binding yet no priming-activity have been achieved in vitro (22), possibly because additional factors required for this rearrangement in HBV but not DHBV ε, are lacking in the current in vitro systems; for instance, binding does not require the apical loop although this sub-element is crucial for priming activity with DHBV. Most importantly, all current information on ε, even for DHBV, was largely restricted to the secondary structure level because the mere size of > 60 nt for either ε element is at the limits of current methodologies for NMR nucleic acid 3D structure determination. We have previously determined the secondary and/or solution 3D structures of two shorter sub-elements (25-28), namely of the PBDs and the apical stem-loops (ALD, ALH in Figure 4.1a) of both human and duck ε elements. While revealing novel structural details, in particular the presence of a tetraloop in DHBV ε and a pseudo-triloop in HBV ε, the investigated RNAs were too much truncated to allow for direct correlations with function; PBD was inactive in in-vitro priming reactions. Generally, all available evidence suggests that P protein recognizes features of the PBD. These features were not adequately modeled by the isolated sub-elements. For instance, elements of the apical stem may be in contact with the primer bulge. Consequently, a 3D model of the complete epsilon allows modeling of such aspects. Moreover, recent studies have shown that dynamics affect specificity and strength of interactions, e.g. fuzzy complex formations (29) or conformational capturing of a particular

RNA conformation (30). Therefore, also the conformational dynamics of the epsilon is of importance.

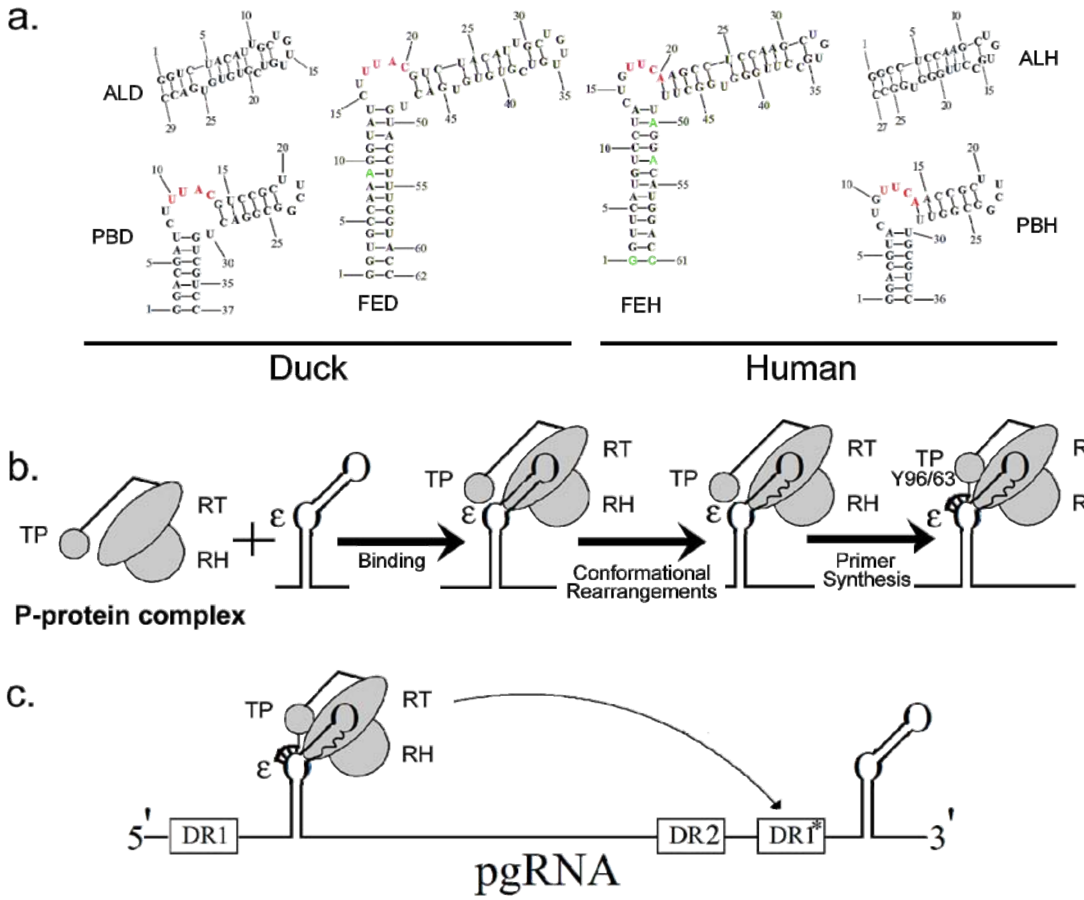


Figure 4.1. a. Secondary structure of the full ~60 nt epsilon of duck (FED) and human (FEH). Also shown are their constituents, namely the primer bulge with the priming sequence written in red (PBD and PBH), and the apical stem loop (ALD and ALH). Some structurally and functionally silent mutations relative to the wild-type have been made in the (full) constructs. All sequences start with a 5'-GG for efficient RNA production. The lower stem of FED is regularized and stabilized by base pairing the wild type bulged U54 to an added A residue (green), leading to A9:U54 basepair. In PBD, the two basepairs underneath the internal bulge are authentic. The remaining part of the lower stem is modified to make it more stable (stack of C4:G34 with G5:C33), while in the upper stem three basepairs above the bulge are maintained as in wild type and the residues and the capping loop in the remaining top part are chosen for high stability and for optimal dispersion in NMR resonances. In FEH, the wild type U9:G53 is mutated to the more stable U9:A53 (green) and wild type U12:G50 to the more stable U12:A50 (green); both mutations are also found in some HBVs (25). In PBH, the 2 basepairs underneath the bulge are authentic as are the 2 basepairs above the bulge; the sequence in the remaining part of the lower stem is chosen for stability and NMR resonance dispersion and similar reasoning governs the sequence and capping of the upper stem. b. The P-protein contains three domains, the RT domain, a RnaseH domain (RH), and a TP-domain. Moreover, a functional P-protein complex requires association of at least two other protein factors (8). After binding of the P-protein complex to epsilon, the RNA undergoes conformational rearrangements, bringing the RNA into a state where primer synthesis is possible. The primer is covalently bound to tyrosine 96 (for duck, and Tyr63 for human) of the P-domain c. After completion of primer synthesis, the primer with its attached proteins translocates to DR1*, from where the (-)-DNA strand of the viral genome is further synthesized.

Here, we present the solution 3D structures, and the dynamics, of the complete ε RNA elements of both the human and duck hepatitis B viruses. This information provides deeper insight into the molecular details of the replication initiation mechanisms of a highly relevant human virus and its most important avian model virus. Clues are obtained for the interpretation of the apparent differences in *in-vitro* functionality of the DHBV versus HBV P-ε complexes, including information on how much of the data derived from DHBV may be transferred one-to-one, or not, to HBV. The data may also provide a rational basis for future functional studies by mutagenesis and information on HBV ε and its interaction with P protein as a potential novel drug target. Finally, the study provides fundamental information on how large RNAs may interact with proteins and information on methodological improvements for their structure determination.

4.3 Results

NMR structure determination The size of both duck and human HBV epsilon elements (~ 62 nt) and the associated resonance overlap in the spectra are challenging for NMR structure determination. Until now only three NMR solution structures of RNAs with a size above 50 nt have been deposited in the PDB. We have used the “divide-and-conquer” approach, introduced by Puglisi and co–workers to derive the structure of the ca. 71 nt IRES RNA element from HCV (31; 32). In this approach, smaller thermodynamically stable sub-elements are excised from the large RNA and studied individually. The structure of the large RNA is then derived from the relative orientation of the known smaller fragments. For the ε RNAs the sub-elements were chosen as displayed in Figure 4.1a.

For the ‘divide-and-conquer’ approach to be valid, the smaller fragments must fold in a conformation similar to that in the context of the large RNA. To validate this assumption, we compared the chemical shifts of nuclei in each of the fragments with those in full-length epsilon (Figure 4.2). The [^{1}H, ^{15}N] HSQC spectra of the imino region of the fragments ALD, FED, and PBD (Figure 4.2 left) show identical imino resonance positions for corresponding basepairs. This demonstrates an identical conformation of the isolated apical stem-loop ALD and the primer bulge PBD or in the context of the full-length epsilon FED. This conformational similarity-is further illustrated by the identical C2-H2 cross peak positions in the [^{1}H, ^{13}C] HSQC spectra (Figure 4.2 right) of corresponding C2-H2 groups in ALD, FED, and PBD (the same conclusion holds for C6-H6 and C8-H8 cross-peak positions in [^{1}H, ^{13}C] HSQC spectra, data not shown). Note that a comparison of the [^{1}H, ^{13}C] HSQC spectra of FED and PBD shows identical C2-H2 (and C6-H6 and C8-H8) cross-peak positions (Figure 4.2 right) for the loop residues of the primer bulge, a similarity that runs into the lower stem and into the upper stem for the two to three residues that are equal (see legend of Figure 4.1). A similar comparison in terms of chemical shifts with equal outcome can be made for the human full-length epsilon, FEH, and its ALH and PBH fragments. Hence, the ‘divide-and-conquer’ approach is applicable to duck and human epsilon and data for the full-length RNAs can faithfully be combined with those for the sub-elements to derive the structures of the full-length ε RNA elements. Below we first describe briefly the structure derivation of the domains, including resonance assignments, and then the structure work on the full epsilon. For duck epsilon, the spectral assignment of the sub-domain PBD was carried out in full detail (33) and its structure and dynamics were determined as described in Supplementary Material. The structure of the ALD sub-domain, and its dynamics, are described in Ampt et al. (34). For human epsilon, the structure and dynamics of PBH was determined as described in Supplementary Material, while the structure of ALH is described by Flodell et al. (27), and its dynamics by Petzold et al. (28).

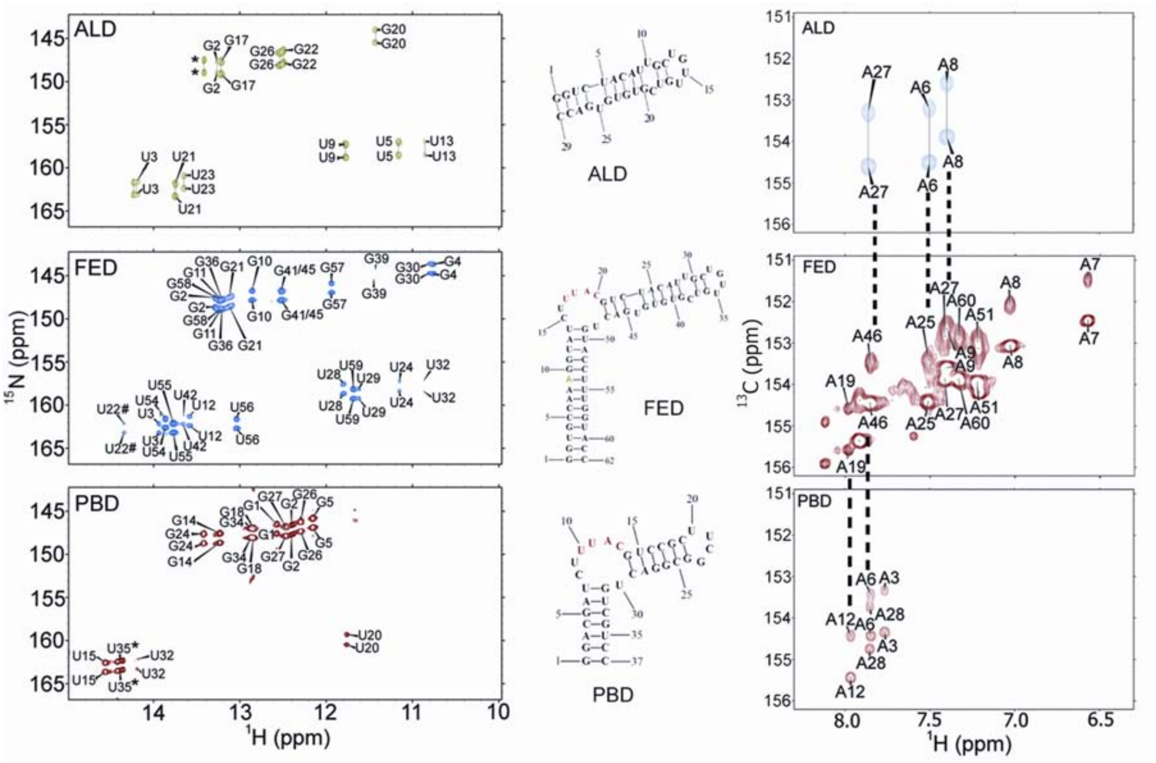


Figure 4.2. Imino [^{1}H,^{15}N] HSQC spectra (15 °C, 600MHz; left panel) and C2-H2 region of [^{1}H,^{13}C] CTHSQC spectra (25 °C, 600MHz; right panel) for ALD, FED and PBD (middle panel), demonstrating the conformational similarity of the fragments (ALD, PBD) in isolated form compared to that in the full-length epsilon (FED). Note that the HSQC were recorded the $^1J_{HN/C}$ present in the indirect dimension (^{15}N and ^{13}C). (Left) Similar cross-peak positions of corresponding imino ^{1}H-^{15}N pairs are observed in the imino HSQC spectra of FED compared to ALD and PBD. Resonance assignments in the imino HSQC of FED follow from the known assignments in ALD and PBD and were confirmed with ^{1}H-^{1}H imino NOESY contacts. Resonances marked with an asterisk (*) show double resonance due to presence or absence of the first basepair in the sequence, which is a consequence of sample synthesis, i.e. U35 in PBD and G2 in ALD. The resonance U22 in FED, marked with a hash (#), does not exactly match the resonance position of its corresponding residues in PBD (U15) and ALD (U3). The difference is caused by slight differences in local structure of FED compared to ALD and PBD. (Right) Equal (similar) cross-peak positions of corresponding H2-C2 pairs are observed in [^{1}H, ^{13}C] CTHSQC spectra (region C2H2) of FED compared to ALD and PBD. The resonances of some corresponding residues are connected via dashed lines for clarity. The resonance assignments of H2-C2 cross peaks in the spectrum of FED translate in part directly from the known assignments of ALD and PBD and were further derived from imino-H2 contacts in the 2D H_2O NOESY spectrum.

For structure calculation of the full-length duck epsilon (FED), a structure calculation protocol was devised that takes account of the domain dynamics. A full description is given in Supplementary Material. Briefly, first, FED starting 3D structures were built by means of fragment replacement using the ALD, PBD, and lower stem A-helix 3D structures. In these starting structures the relative orientation of the upper and lower stems still need to be optimized as well as the structure of the primer bulge. The orientation of the upper stem relative to the lower stem in FED was derived from a grid search based on FED RDC data using the gyration tensor method (35) to calculate the alignment tensor of the test structures and minimizing the RMSD between calculated and experimental RDCs. To account for domain and primer bulge dynamics the following approach was taken. First, an ensemble of structures was retained with an RMSD_{RDC} up to a size of 3σ, where σ stands for the experimental error in the measured RDCs. It was established that the half angle of the cone describing the relative helix orientations of this set of structures equals the cone half angle derived for domain dynamics from NMR relaxation data. In this initial set of 3D structures that correctly describes the helix orientations and their relative motion, a correct primer bulge structure was not yet a requirement. Second, a molecular dynamics refinement protocol was then carried out on each member of this ensemble. In the protocol, the global structure, that is the conformation and relative orientation of the upper and lower helices, were retained, while the conformation of the primer bulge was refined against RDC data (from FED as well as from PBD) and other NMR data as well as holonomic restraints. In this way, an ensemble of structures was obtained that is fully compatible with all NMR and holonomic restraints. A similar approach was taken for deriving the structure of the human epsilon, except that dynamics were not considered (Supplementary Material).

Structure of Duck and Human Epsilons

Overall structure The two epsilon-domain 3D structures are strikingly similar in their global conformation. Both are fairly straight with the primer bulges inducing on average angles of ~70° (HBV) and ~50° (DHBV) between the lower and upper stems (Figure 4.3b). In the apical stem–loop, we find in the human variant that the bulged U43 is partly intercalated resulting in a bend of ~20° in the apical stem, while in the duck variant the corresponding bulged U44 is almost completely excluded from the helical stack, resulting in a smaller bend of ~10°. Nevertheless, an overlay of the entire stem-loop (Figure 4.3b) shows that the overall structures largely overlap. However, differences are seen in the primer bulge.

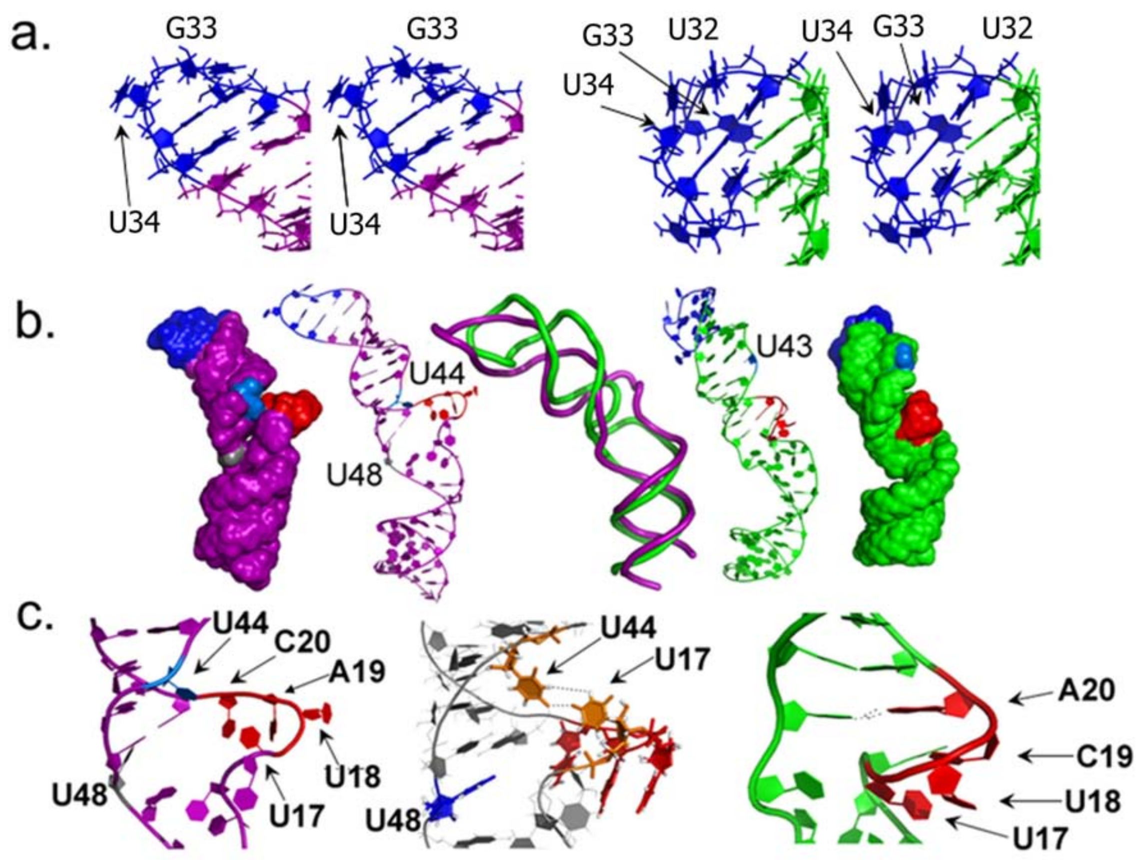


Figure 4.3. Overview and details of the human and duck epsilon NMR solution structure. a. Stereo displays of the UGUU tetraloop of the FED (left) and the pseudotriloop of FEH (right) b. FED on the left, and FEH on the right. From outer to inner structures, the surface of the structures is displayed, with the primer sequence marked in red, the bulged out residue in the apical stem in marine (U44 in FED, and U43 in FEH), and the loops shown in a are marked in blue. In FED the inserted residue in the strand, opposing the primer bulge is marked in grey. The same color-coding holds for the cartoon display. In the middle, an overlay of the backbones of FEH and FED is displayed. c. From left to right are displayed two conformers of the primer bulge of the duck epsilon, and to the far right the primer-binding region of the human epsilon.

Primer bulge and apical loop structures In human ε, residue A20, which is the 3' residue of the UUCA primer template sequence (Figure 4.1a), is part of the apical stem and base-pairs with U48 (Figure 4.1a, 4.3c). In contrast, the corresponding residue C20 in duck ε, which is the 3' residue of the UUAC primer template sequence (Figure 1a), is part of the primer bulge (Figure 4.1a, 4.3c). In human ε, three bulge residues (C14-G16, Figure 4.1a) precede the UUCA primer template sequence, while in duck ε two residues (C15 and U16,-Figure 4.1a) precede the UUAC primer template. This has interesting consequences for the three-dimensional structure of the two epsilons. Instead of inducing an extra rotation, the presence of this extra residue in human ε is compensated for by a change in conformation of the bulge, so that the apical stems of human and duck ε remain largely in overlap when overlaying the lower ε stems (Figure 4.3b). As a result, the apical UGUU tetraloop in DHBV and the UGU triloop of the pseudotriloop of HBV point essentially in the same direction when their lower ε stems are overlaid (Figure 4.3b).

At the tip of the UGUU apical loop of duck ε (Figure 4.3a), residue G33 stacks more or less onto U32, which forms a U:U base pair with U35, while U34 points into the solution and folds somewhat back into the minor groove. In a similar fashion, at the tip of the UGU triloop of human ε (Figure 4.3a), U32 stacks onto the underlying C:G base pair, while G33 folds in one conformation - shown in Figure 4.3a - back into the minor groove; U34 points into the solution and/or stacks with U32. In human ε, the bulged C36 of the pseudotriloop is fully solvent exposed.

In duck ε, the bulged residue U44 lies outside the stack of the apical stem and points into the primer bulge, where it can occasionally form hydrogen bonds with U17 as evident from the two members of the ensemble of structures (Figure 4.3c, left and middle). This is an example of an interaction that would not at all be envisaged based only on 2D information, and emphasizes the novel information that can be gained from full three-dimensional structure determination. In contrast, in human ε, the bulged U43 lies far away from the primer bulge. This difference is caused by the number of basepairs between the bulge U and the primer bulge, namely 3 basepairs in duck ε and 5 basepairs in human ε.

A further significant difference between the two ε RNAs is the region of the primer bulge. In duck ε, an extra-unpaired residue (U48) is present in the strand opposite the UUAC/UUCA template strand of the primer bulge loop. A way to view this presence is to note that in human ε, U48 has become base paired with A20, the 3’-residue of the primer template sequence UUCA (Figure 4.3c, right). This base pairing pulls the backbone inward (Figure 4.3c, right), so that in human ε the residues of the primer bulge reside in the locally widened major groove (Figure 4.3b). In duck ε, this base pairing is absent. Consequently, going into the primer bulge region from C20 back to C15, the backbone (C20-U18) continues to follow more or less the helical trajectory of the apical stem. As a result the 3’-proximal part of the bulge (C20—U17) is quite solvent exposed (Figure 4.3c, left). After this initial trajectory into solution, the backbone makes a sharp turn at U18 (Figure 4.3c, middle) to be able to connect to the lower stem.

In both duck and human ε, the primer bulge can be viewed as having an S-shaped conformation. In human ε, the 5’-proximal part of the bulge (C14-G16) is fully single stranded and lies in the widened major grove, while its 3’-proximal part (U17-A20) is only partially single-stranded due to the A20:U48 basepair. In duck ε, the 3´-proximal part of the bulge is instead fully single-stranded due to the absence of a C20:U48 basepair, while its 5’-proximal part is shortened, because the bulge lacks here the extra G present in human. Finally, the primer template residues are fully solvent exposed in duck ε, but only partially so in human ε.

The three basepairs underneath the primer bulge in the lower stem are essential for productive binding, i.e. for priming competence (9). They are located away from the primer bulge residues described above. Interestingly, the residues in the two base pairs closest to the primer bulge display considerable degrees of internal dynamics, more so then the corresponding base pairs in the apical stem (see below).

Epsilon internal motion

Biomolecules generally do not interact with their partners as rigid bodies. Rather, their ensembles of structures change adaptively to optimize intermolecular interactions through the well-known ‘induced-fit’ mechanism and/or via the recently proposed and established ‘conformational selection’ mechanism (36; 37). ‘Conformational selection’ postulates that all biomolecular conformations pre-exist and the ligand selects the most favored conformation. Following binding the ensemble undergoes a population shift, redistributing the conformational states. Both conformational selection and induced fit appear to play important roles, i.e. following binding by a primary conformational selection event, optimization of interactions is likely to proceed by an induced fit mechanism (36; 37). For instance, conformational selection was shown to be the mechanism by which in the protein world, ubiquitin recognizes its many different protein partners (38). The structural and dynamic diversity of RNA rivals that of proteins (39). Adaptive recognition is particularly common in flexible RNAs where conformational changes provide the basis for sensing, processing, and transmitting chemical information (30; 36; 39; 40). For instance, HIV TAR RNA was shown to bind its peptide ligand via a conformational selection mechanism. Conformational selection has now been observed for protein-ligand, protein-protein, protein-DNA, protein-RNA and RNA-ligand interactions; thus, conformational selection appears to be universal in molecular recognition processes involving biomolecules (37). It has thus been proposed that these data support a new molecular recognition paradigm for processes as diverse as signaling, catalysis, gene regulation and protein aggregation in disease, which has the potential to significantly impact our views and strategies in drug design, biomolecular engineering and molecular evolution (37). We therefore investigated the internal motion of epsilon and its domains on a variety of time-scales. Previously, NMR relaxation data were acquired and analyzed for ALH (28) and ALD (26). Here, we present the NMR relaxation data on PBD and indicate how this information can be translated into the dynamics for the full epsilon (FED).

For PBD, the generalized order parameters S^2 of C-H bond vectors were derived from the NMR relaxation data and are displayed in Figure 4.4. The generalized order parameters S^2 are indicative of internal motion on the ps to ns timescale. They describe the degree of order of the C-H bond vectors within the molecular frame. A value of 1 indicates a completely ordered C-H bond vector and thus a rigid vector, while a value of 0 indicates a completely disordered C-H vector and thus a completely mobile C-H vector within the molecular frame. Modeling the motion of the C-H vector as free diffusion in a cone relates the cone's opening angle θ with S^2 and thus provides a view of the degree of motion ($S^2 = \cos(\theta)(1+\cos(\theta))^2$; $S^2 = 1$, $\theta = 0°$; $S^2 = 0.8$, $\theta = 22°$; $S^2 = 0.6$, $\theta = 32°$; $S^2 = 0.4$, $\theta = 42°$; $S^2 = 0.2$, $\theta = 55°$). As can be seen in Figure 4.3 from the reduced S^2 values, the primer bulge residues experience considerable motion within the molecular frame with S^2 reaching as low as 0.2 ($\theta = 55°$). Note also that the juxtaposed hinge residue U30 in PBD (U48 in FED) is quite mobile as evident from its reduced S^2. Consequently, the rigid upper and lower helices are connected via a flexible hinge in both strands introducing domain motions of the helical stems with respect to each other. Because FED and PBD have these primer bulge regions in common, these same motions will thus be present in the FED.

The degree of domain motion can be qualitatively estimated from a comparison of the S^2 values of PDB and ALD (see also Figure 4.4). Such a comparison shows that the primer-loop indeed induces extra helix-stem-domain motions in PBD as compared to ALD. Specifically, comparison of the S^2 of the PBD upper helix and with S^2 of the ALD helix-stem shows that within PBD the S^2 is lower, indicating additional domain motion, described by an $S^2_{\mathrm{domainPBD}}$ of 0.80(7). A similar slightly higher estimate of $S^2_{\mathrm{domainPBD}}$ is obtained (0.85(6)) when comparing the squared order parameter for ns time scale motion, S^2_{s}, of PBD and ALD derived from the NMR relaxation data (see legend Figure 4.4). The domain motion $S^2_{\mathrm{domainPBD}}$ value of 0.80(7) corresponds to a cone with half opening angle of 21(.4)°, assuming the wobbling in a cone model for the domain motion (41).

In Figure 4.4, the S^2 data from PBD and ALD were combined to provide a view of the S^2 values within the full duck epsilon FED. In summary, within FED the residues in the primer bulge and the loop capping the apical stem display considerable intra-molecular motion as evident from the lower S^2 of these residues. In the capping tetra loop (UGUU), in particular the solvent-exposed residue U34 shows increased intra-molecular motion. In the primer bulge region, a significant degree of intra-molecular motion is observed for residues 15 to 20 and for residue U48 in the opposing strand (arrowed residue in Figure 4.4). Because of this local dynamics in the primer bulge, the helix domain motion can take place.

In the structure calculations of FED this domain dynamics as well as the local dynamics in the primer bulge and apical loop were accounted for as much as possible in constructing the structure ensemble (see above and Supplementary Material). The ensemble of resulting FED structures is shown in Figure 4.4. They are displayed in such a way that the lower stem is overlaid, so that the trajectory of domain motion is well visualized. To establish how well the ensemble of structures describes the experimental S^2 values, we calculated these values from the final ensemble (Supplementary Material). The domain motion is accounted for because the upper stem wobbles in a cone with a half-angle of 21° corresponding to a back-calculated S^2 of ca 0.8. Moreover, the smaller S^2 values experimentally derived from NMR relaxation data for the primer bulge and apical tetra loop are semi-quantitatively represented in the ensemble of NMR structures. In other words, in the ensemble of duck ε structures, the intra-molecular domain motions as well as the local intra-molecular motion in the primer bulge and apical loop are semi-quantitatively accounted for.

4.4 Discussion

Human and duck epsilon their solution structures and their dynamics

The three-dimensional solution structures of human and duck ε are remarkably similar both displaying a rather elongated shape with bend angles of only ~50° and ~70°, respectively (Figure 4.3b). An elongated shape as seen here in human and duck ε appears of biological significance, because upon initial binding ε should fit into the palm of the hand formed by the RT domain (21) and this requires an elongated shape. In addition, it is evident from the NMR data that the three- dimensional structure of duck ε experiences a variety of internal motions of different sorts that appear functionally relevant.

First, a relatively small but still significant amount of domain motion is superimposed onto the three-dimensional structure (Figure 4.4). The upper and lower helices undergo domain motion of a degree that can be estimated to have a squared order parameter S^2 of ca. 0.80, which corresponds to a cone half angle of ~20° (Figure 4.4). This domain motion is likely caused by the heightened degree of internal motion observed in the primer bulge (Figure 4.4), and is likely advantageous for the initial binding of ε to the RT domain as it allows for small adjustments or selection of correct elongated shape to fit into the palm of RT hand. A smaller or similar degree of domain is expected to be present in human ε. Second, a considerable degree of internal motion is observed in the tip of the apical UGUU tetraloop (26), where the highly mobile loop residue U_{34} (3rd loop residue) is essential for binding RT. This again points to the functional importance of dynamics. We note in passing that the apical pseudo tri-loop of human ε also displays considerable internal motion (28).

Third, the middle part of the apical stem of duck ε is thermo-sensitive and appears to melt at 35 °C(26). NMR relaxation measurements have shown that this instability leads, not to additional internal motion on the ps- to ns-time scale, but to additional base-pair opening (26). Stabilization of this stem region abolishes RT binding (9). In other words, this base-pair opening is essential for binding. It thus appears again that both the three-dimensional structure and its internal motion are important for binding.

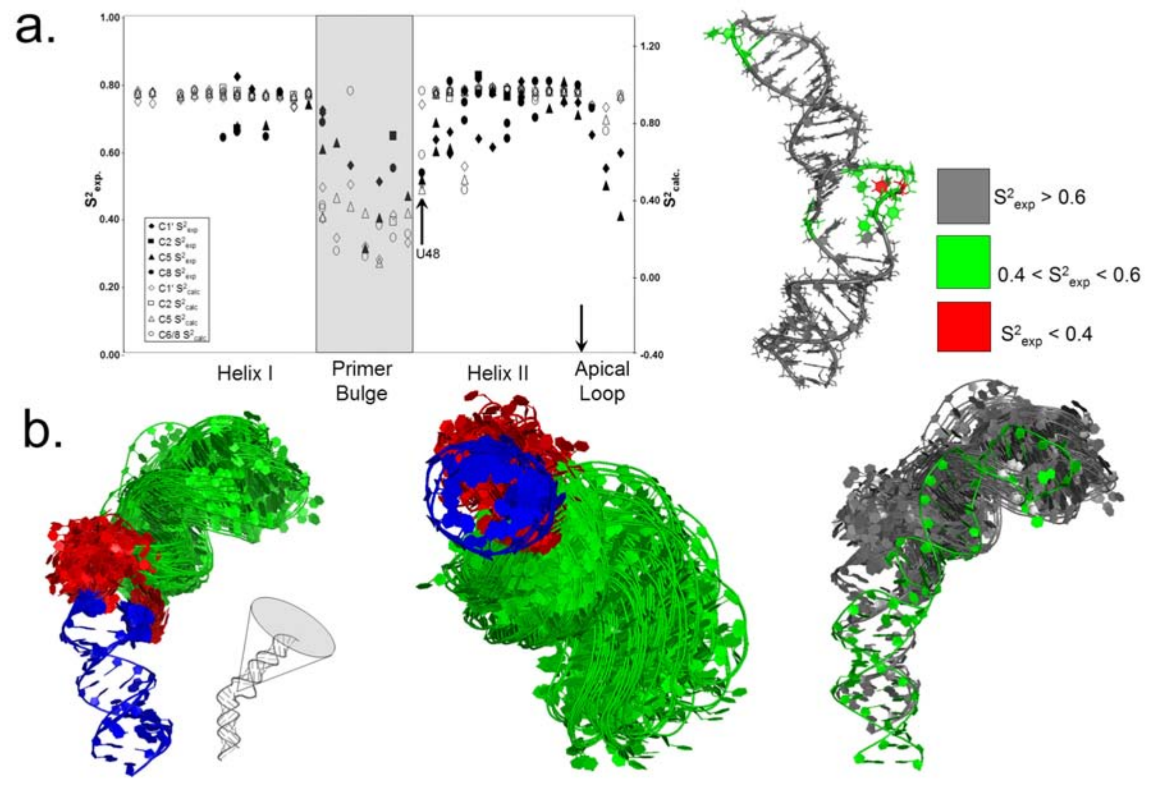


Figure 4.4. Structure and dynamics of duck epsilon. a, left. Squared order parameters, observed and back-calculated from the ensemble of FED structures. The meaning of the data-point symbols in the panel is as given in the boxed legend in the left bottom. The observed data are displayed as follows. The squared order parameters of PDB, S^2_{PBD}, are displayed as measured (Supplementary Material). For visualization purposes and comparison with FED, the S^2 of ALD, S^2_{ALD}, was multiplied (scaled) with the estimated domain motion, $S^2_{domainPBD}$, $S^2_{ALD*,FEDapicalstemloop} = S^2_{ALD} \times S^2_{domainPBD}$. The domain motion, $S^2_{domainPBD}$, was estimated as follows from the NMR relaxation data. From the NMR relaxation data it is evident that on average the S^2 of the helical stems of PBD is lower than that of the apical stem of ALD (26) and/or of ALH (28): PBD, $<S^2_{lower/upper\ stem,\ PBD}> = 0.70(4)$; ALD, $<S^2_{stem,\ ALD}> = 0.85(7)$; ALH, $<S^2_{stem,\ ALH}> = 0.89(9)$. This demonstrates the presence of domain motion in PBD, and thus in FED, a motion induced by the primer bulge. The squared order parameter of the stem of PBD, $S^2_{stemPBD}$, may thus be written as the product of the squared order parameter for fast internal motion, $S^2_{f,stemPBD}$, and for domain motion, $S^2_{domainPBD}$, $S^2_{stemPBD} = S^2_{f,stemPBD} \times S^2_{domainPBD}$. The fast internal motion in the stem of PDB may be assumed equal to the usual A-helix stem values and thus estimated equal to the average of $<S^2_{stem,\ ALD}>$ and $<S^2_{stem,ALH}>$. The domain motion in PBD, $S^2_{domain\ PBD}$, estimate is then $S^2_{domainPBD} =< S^2_{stemPBD} > /(0.5(< S^2_{stemALD} > + < S^2_{stemALH} >))$; $S^2_{domainPBD}$ is then 0.80(7). Alternatively, $S^2_{domainPBD}$ can be estimated from S^2_s obtained from relaxation data analysis; $<S^2_{s,\ PBD}>$ equals 0.76(4) for helix residues. However, this low $S^2_{s,\ PBD}$ value cannot fully be attributed to helix domain motion, because A-helices may display also some slow (ns time scale) motions as evident from the NMR relaxation analysis of ALD and ALH, whose residues do sometimes display some slow (ns) motion. In ALD and ALH, S^2_s equals 0.87(6) and 0.90(8), respectively. The $<S^2_{s,\ PBD}>$ of 0.76(4) for helix residues is indeed lower than S^2_s of ALD and ALH. Correcting for S^2_s in A-helices by dividing $<S^2_{s,\ PBD}>$ by $<S^2_s>$ from ALD/ALH, one finds the second estimate, $S^2_{domainPBD2}$ is ca. 0.85(6), a number quite close to that of earlier estimate. We chose to use $S^2_{domainPBD}$ to correct S^2_{ALD} for domain motion because its derivation is more direct. a, right. Visualization of the dynamics in the 3D structure of full-length epsilon, FED, by way of color coding of the residues: highly mobile ($S^2 < 0.4$), red; intermediate degrees of mobility ($0.4 < S^2 < 0.6$), green; more rigid ($S^2 > 0.6$), grey. b. The ensemble of FED structures is displayed keeping the bottom (blue) helix aligned; in this way the trajectory of helix domain motions is visualized (figures left and middle). >>

The apical stem (green) moves in a cone of ~19°, displayed schematically in gray below the left figure. Note that in solution both helices move; here, one helix is fixed for display purposes. The lowest energy structure is shown in green in the right figure, superimposed on the entire ensemble of structures that cover the cone in which motion takes place.

Note that based on mutation data it has been proposed that formation of the priming-active epsilon-RT complex is a two-step process (9), i.e. initial recognition and binding of the free RNA by the RT (step 1) is followed by major conformational rearrangements required for and leading up to formation of the priming-active complex (step 2). If we accept a two-step process then the current data are highly relevant for the initial specific binding. This is what the RT gets to see of the RNA and based on this, binding occurs or not. Although, no direct information on the structure of the RNA in the priming-active conformation is available, from the now-known 3D structure of the free RNA we can also make some reasonable predictions on some aspects of the priming-active structure (as discussed below). Taken together, the 3D structure of the free RNA poses new constraints, which allows for predictions which then could be tested by functional mutagenesis. Of course, on a fundamental note the value of the now-known 3D structure of the free ε RNAs lies in the fact that to demonstrate a conformational change, one needs to know at least two different structures of which in this case that of the free RNA is one. Against this 3D structure the new conformation can later be compared.

Below we will discuss the three-dimensional structure and dynamics of ε in the context of similar structural RNA elements followed by a discussion of the correlation between mutational data and ε three-dimensional structure and dynamics, and finally the structural consequences for the initial ε-RT binding and the ultimate priming active ε-RT complex.

The duck and human ε belong to a rare 6-nt bulge family

Structurally, the highly conserved human and duck HBV ε elements can be considered to belong to the family of asymmetric 6-nt internal bulge loops; specifically, human and duck HBV ε elements belong to the family of asymmetric 6-nt internal bulge loops with lengths of (6-nt, 0-nt) and (6-nt, 1-nt), respectively, where x-nt and y-nt in (x-nt, y-nt) stand for the numbers of bulged residues in the connecting strands. Up to 4-nt, asymmetric internal bulge loops are quite abundant structures (42). Asymmetric internal bulge loops of lengths 3-nt (e.g.

TAR-HIV) and 4-nt have been found (43) or predicted (42) to be quite flexible. On the other hand, asymmetric 5-nt bulge loops appear relatively rare occurrences (44) with only two examples (31; 44). One of them is an asymmetric 5-nt internal bulge (J2a/b) in the human telomerase RNA core region that is highly conserved in its location but not sequence; the internal bulge loop forms a defined S-shape and creates an ~90 ° bend with a surprisingly low twist (~10 °) between the flanking helices (44). The other example of a 5-nt bulge is found in the hepatitis C virus internal ribosome entry site (HCV IRES) and has a different sequence but similar structure (31). The 5-nt internal bulge loops are remarkably stiff in comparison with their shorter 3/4-nt counterparts. For instance, the cone half angle for domain motion of the 5-nt bulge J2a/b is ~19° (44), while a value of ~60° applies for the 4-nt asymmetric bulge loops (42).

Asymmetric internal bulge loops of 6-nt length appear also rare. Apart from the duck and human HBV ε structures reported here by us, we found only one example using RNA FRABASE (45), namely the asymmetric (6-nt,0-nt) internal bulge loop in the loop B RNA of the stem-loop IV domain of the internal ribosomal entry site (IRES) of entero viruses (EV) (46). The asymmetric 6-nt internal bulge loop of EV-IRES folds in solution into an L-shaped (bend angle ~90 °) conformation as well as a U-shaped (bend angle >> 90 °) structure depending on the sequence. The duck and human HBV ε display, like the L-shaped EV-IRES, bend 3D structures (Figure 4.3b), although they are a bit more elongated, that is oblique L-shaped with bend angles of 70 ° to 50 °. The residues in the internal bulge loop of the U-shaped structures show very little dynamics, while those of the L-shaped structure display considerable mobility. From the NMR relaxation data on duck HBV ε and its separate domains, we can conclude that the asymmetric 6-nt bulge loop residues of duck HBV ε are quite mobile like the ones in the L-shaped EV-IRES. Importantly, the domain motion of the flanking helices could be estimated from the NMR relaxation data and corresponds to a cone half angle of ~20°, a value quite similar to the ~19° domain cone motion seen for the telomerase RNA core domain (44). Hence, like 5-nt bulges, the domain motion of 6-nt bulges is remarkably small compared to that of 4-nt bulges. In conclusion, the asymmetric 6-nt internal bulge loops of human and duck HBV ε show considerable similarities in terms of 3D structure and dynamics to the asymmetric 6-nt internal bulge loop of EV-IRES as well as to the asymmetric 5-nt bulge loops of the HCV IRES element and human telomerase RNA core domain.

The 5-nt bulge in HCV IRES (domain II) has a completely different sequence than J2a/b, but the structure is remarkably similar (31). It has been proposed that the bend in domain II is a conserved structural feature of HCV-like IRES elements that relates to a structurally but not a sequence encoded function (47). Indeed, mutations within domain II that affect its overall shape also impair function (47). In telomerase, J2a/b is also not conserved in sequence but is conserved in location and was therefore proposed to be a structurally encoded functional domain (44). Because telomerase is highly active in most cancer cell lines but not in somatic cells, it has attracted wide interest as a potential drug target. It was therefore argued that the unusual features of J2a/b make it an attractive target for small molecules, i.e. small molecules could bind the bulge and thus affect the bend angle and/or dynamics and thereby affect telomerase function (44). While results for telomerase are not yet available, similar reasoning has for HCV already led to the successful development of small molecule inhibitors of HCV, inhibitors that target the 5-nt bulge in HCV IRES (48; 49). In a similar vain, the 6-nt bulge of duck and human HBV ε can be considered a structurally encoded functional domain, as the conserved elongated (oblique L-shaped) shape of human and duck HBV ε is important for initial binding to the RT domain. Hence, the 6-nt bulge of duck and human HBV ε could form an interesting target for small molecule inhibitors which function by affecting the bulge conformation and/or dynamics. For instance, inducing a U-shaped conformation prevents binding into the palm of the hand of the RT domain. Alternatively, rigidifying the 3D structure of ε via the 6-nt bulge could prevent affective binding and subsequent primer synthesis.

Correlation between mutations and structure and dynamics

SELEX studies (9) and additional recent functional studies by some of us (Nassal lab) have shown that in duck ε the upper stem sequence is not really important for function, but instead stability or lack of stability of the upper stem. Various mutants, like those described in the *in-vitro* SELEX paper (9), are functional in infection, although none is as good as the wild-type. As long as the major requirement is met, namely no stable base pairing in upper stem, the length of the upper stem can be modified (+4 nt, -4 or even more nt's) without completely preventing formation of a priming-active complex *in-vitro*. Some of these mutants are even infectious in ducks. The conclusion is therefore that the upper stem sequence and length are not essential for binding AND rearrangement. In contrast, the apical tetra loop sequence IS important. This is evident form the following observations. a) In-vivo, a U35G mutation in the loop is not genetically stable but reverts back to wild-type U. b) Further, if the complete "classical heptaloop" (C<u>UGUU</u>GU, tetraloop underlined) sequence is randomized in the context of the so-called S1-mutant again (see reference (9)), and the corresponding viruses are inoculated into ducks, a rapid selection of one sequence takes place that is g<u>aGUU</u>GU. Hence, 5 out of 7 positions are as in wild-type. Whether the 5´-ga indicates an influence of the S1 context is not yet clear. A virus pool with randomized heptaloop sequence in an otherwise wt-context is presently generated for inoculation into ducks. However, the experiment will, unfortunately, take a while to complete. c) Mutation of the solvent-exposed loop residue U34 to C34 completely abrogates binding.

The U48G mutation abolishes priming but not binding (9). Interestingly, in all conformations of the flexible primer bulge U48 is pointing into the primer bulge stack and in most conformations C20 points toward U48. Consequently, the present structure suggests that a C20:G48 base pair is likely to form in the U48G mutant, thereby lengthening the upper stem with a highly stable C20:G48//G21:C47 base stack. After initial binding, major conformational changes occur to form a priming-competent complex. Thus, stabilization of the bottom of the upper stem appears not to abolish the binding, but instead, the mutations in and around the primer bulge, e.g. like U48G, primarily affect the priming. Consistently, productive (binding and priming) depends also on the nature of upper two base pairs in the lower stem, U14:G49 and A14:U50. In particular, residue G49 is found to be essential, but not its base pairing (24). Interestingly, we find that G49 is quite mobile and base pairs only part of the time. The same high degree of internal motion is observed for the residues of the next putative base pair, A13 and U50. One might thus predict that it is the nature of the residue which is important and not the base pairing. Further, the major conformational changes and unfolding that appear to occur on the way to a priming-competent complex, suggest G49 (and possible U50) directly interact with the RT protein.

We find in the complete duck epsilon the whole plethora of types of internal motions, i.e. tertiary, secondary, and loop dynamics occurs. This suggests that all these aspects of dynamics are important if not essential for interaction with the RT. For instance, in duck, the apical stem-loop of epsilon displays, within its weak stem, enhanced basepair opening and thus enhanced internal motions (26). Stabilization of the stem abolishes binding to the RT domain, demonstrating that these motions are essential for binding (see above). The interaction mechanism thus likely involves conformational selection within a secondary structure rather than a tertiary structure. In human, the apical stem-loop the stem is highly stable, which in this case does not abolish RT binding.

Model of the initial ε-RT complex and some implications for priming-active complex

Nassal and coworkers (7-9; 23) have demonstrated from extensive mutation studies on *in-vitro* duck HBV systems that the ε-RT interaction is a two-step process (Figure 4.1b). Initially, ε recognizes and binds to the RT domain without major conformational change. With the 3D structures of free human and duck HBV ε and reliable 3D homology models of the HBV RT domain (21) in hand, it is then possible to make a reasonable model of the initial ε-RT complex. Such a model may allow some conclusions for future work to be drawn. Note that after this initial binding, major conformational rearrangements take place that lead up to the priming-competent complex. A schematic drawing of the model of the initial ε-RT complex is shown in Figure 4.5. It shows that placing the 4-nt primer template sequence UUAC over the active site in the RT domain leads to a conformation in which the apical loop of ε is shifted out of the palm of the RT domain. In this way, the apical loop does not interact with the RT domain. Such a model is not correct, because strong evidence exists that in duck the apical loop interacts with the RT domain upon initial binding. 1) In duck, it has been shown that the presence of the apical loop is essential for binding and that within the apical loop residue U34 is essential for binding (9). 2) The loop requirement holds even with a chaperone-independent mini-polymerase construct, miniRT2 (50; 51) or miniPol (own work, M. Nassal); the observation that priming is achieved with the miniRT (50) or the miniPol (own work, M. Nassal) as the only protein, that is with no other factors (capsid, cell, chaperone protein Hsp) present to which the apical loop could bind, implies that binding must be to the RT domain. Hence, the apical loop and specifically residue U34 must therefore interact with the RT domain, most likely with the outer edge of the palm of RT. The implication is of this apical loop/U34-RT interaction is that the primer template sequence moves at the other end of the palm out of the RT domain beyond the active of the RT domain (Figure 4.5). When keeping this U34-RT interaction in place, the primer template sequence can only approach the RT active site after conformational rearrangement, i.e. melting of the apical stem.

Interestingly, for human HBV, an *in-vitro* model for ε-RT interactions was also developed; however, it only shows ε-RT binding, a priming-active complex is not formed (52; 53). Removal of the apical loop in human ε does not inhibit binding, in contrast to the situation in duck ε, but the apical loop is essential for encapsidation (52). Thus, it appears that in human, the apical pseudo-triloop does not interact with the RT domain. It is then free to interact with the (capsid) proteins like pseudo-triloops are known to do in other cases (54) or with as yet unknown cell factors. This interaction may provide the energy to overcome the high stability (T_m = 80 °C) of the base pairing in the human apical stem (10), because unfolding of the apical stem is needed to form a priming competent complex. Interestingly, in duck it is essential for priming that the apical stem has a relatively low stability and its mid-segment melts already at ca. 37°C (10). Furthermore, lengthening of the instable stem does not prohibit priming (own work, M. Nassal).

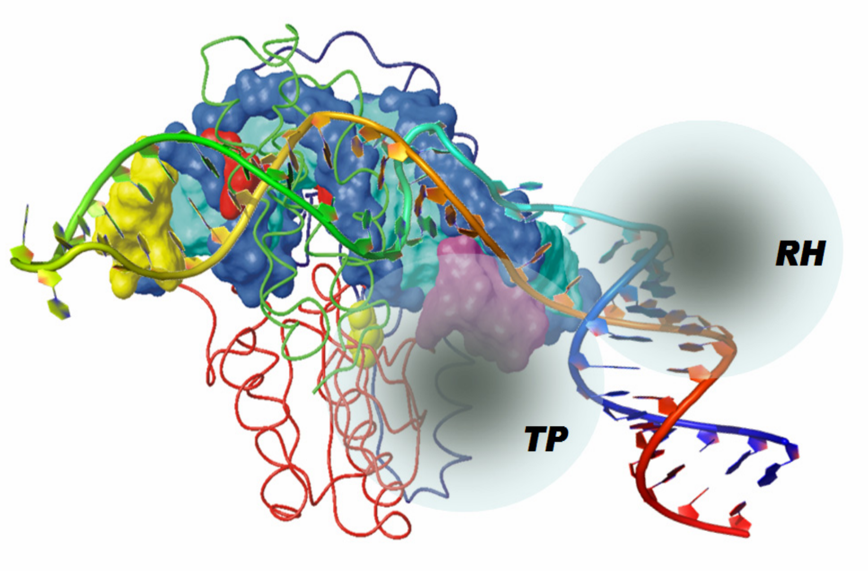


Figure 4.5. The 3D structures of the free duck and human HBV ε, FED and FEH, respectively, docked into the palm of the 3D model of the human HBV RT-domain. The 3D model of the human HBV RT-domain was homology modeled by Das et al.(21) based on the crystal structure of the HIV RT domain. The red and blue wireframe sections show the thumb and palm of the hand of the RT-domain. The FEH is shown in surface presentation (surface representation, truncated below the primer region, purple). It is docked into the hand, and the primer region (purple) is located over the palm of the edge of the palm of the hand near the active site. >>

In cartoon representation, the FED is superimposed, showing that the primer loop is also located at the active edge of the palm of the hand of the RT domain. The two semi-transparent spheres schematically indicate the location of the RH and TP domains of the P-protein complex.

4.5 Concluding Remarks

The 3D structures of human and duck HBV ε belong to the class of asymmetric 6-nt internal bulge loops. In contrast to 1-nt to 4-nt bulges, 5-nt and 6-nt asymmetric internal bulge loops form rare occurrences in nature. The 6-nt bulges of human and duck HBV ε show considerable similarities in terms of 3D structure and dynamics to the asymmetric 6-nt internal bulge loop of EV-IRES as well as to the asymmetric 5-nt bulge loops of the HCV IRES element and human telomerase RNA core domain. Like 5-nt bulges, the domain motion of 6-nt bulges as seen in ε is remarkably small compared to that of 3/4-nt bulges (HIV). Like the 5-nt bulges of telomerase and HCV, the 6-nt bulge of duck and human HBV ε can be considered structurally encoded functional domains and could form an interesting target for small molecule inhibitors which function by affecting the bulge conformation and/or dynamics. Such inhibitors have already been developed for HCV using the 5-nt bulge as target. The 3D structures of duck and human HBV ε presented here may thus provide a framework to help the design new molecules that prevent the crucial RT interaction, i.e. new anti-virals. They can also help to design more insightful mutations to further unravel the RT interaction. Evidently, the epsilon structures are dynamic and do not lock into a single conformation. One has to view the RT-epsilon interaction therefore as a dynamic process. The insights on epsilon-RT interaction obtained from the duck epsilon structure and biochemical data can be translated to human epsilon-RT interaction. Although, overall quite similar, the present 3D structures also show significant differences. For instance, the 3D structures showed that in duck ε the bulged U in the apical stem interacts with residues in the primer bulge loop, while such interactions are absent in human ε. This observation could not have been made without the aid of the 3D structure. The extra base pairing in the primer bulge of duck ε likely stabilizes the primer bulge loop of duck with respect to that of human ε and thereby reduces domain motion of the helices flanking the primer bulge. Another such example, comes from the 3D structure of human ε and is the formation of an extra A:U basepair at the bottom of the apical stem involving the 5’-residue A of the primer template and the U residue in the opposite strand; comparison with the 3D structure of duck ε shows that this basepair results from a change in the primer template sequence, namely from UUAC in duck to UUCA in human. Formation of this A:U basepair likely stabilizes the primer bulge region of human ε relative to duck ε and thereby reduces domain motion of the helices flanking the primer bulge loop. These observations demonstrate the presence of a balancing of stabilizing factors between duck and human ε; they may explain the relatively small degree of domain motion observed in the 6-nt bulge loop compared 3 to 4-nt bulge loops. These and other observations from the 3D structures and the dynamics may help to further our understanding of the functional features.

Acknowledgements

The 6th framework program of the EU, project FSG-V-RNA, is acknowledged for funding.

4.6 References and notes

1. Summers J, Mason WS. 1982. Replication of the Genome of a Hepatitis B-Like Virus by Reverse Transcription of an RNA Intermediate. *Cell* 29:403-15
2. Lee WM. 1997. Hepatitis B Virus Infection. *N Engl J Med* 337:1733-45
3. Hilleman MR. 2003. Critical overview and outlook: pathogenesis, prevention, and treatment of hepatitis and hepatocarcinoma caused by hepatitis B virus. *Vaccine* 21:4626-49
4. Kidd-Ljunggren K, Holmberg A, Blackberg J, Lindquist B. 2006. High levels of hepatitis B virus DNA in body fluids from chronic carriers. *J.Hosp.Infect.* 64:352-7
5. Ganem D, Prince AM. 2004. Hepatitis B Virus Infection - Natural History and Clinical Consequences. *N Engl J Med* 350:1118-29
6. Nassal M. 2009. New insights into HBV replication: new opportunities for improved therapies. *Future Virol.* 4:55-70
7. Nassal M. 2008. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* 134:235-49
8. Beck J, Nassal M. 2007. Hepatitis B virus replication. *World J Gastroenterol* 13:48-64
9. Hu K, Beck J, Nassal M. 2004. SELEX-derived aptamers of the duck hepatitis B virus RNA encapsidation signal distinguish critical and non-critical residues for productive initiation of reverse transcription. *Nucleic Acids Res* 32:4377-89
10. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35:2800-11
11. Schultz U, Grgacic E, Nassal M. 2004. Duck hepatitis B virus: An invaluable model system for HBV infection. In *Adv. Virus Res.*, 63:1-70. San Diego: Elsevier Academic Press Inc. Number of 1-70 pp.

12. Wang GH, Seeger C. 1992. The Reverse Transcriptase of Hepatitis B Virus Acts As a Protein Primer for Viral DNA Synthesis. *Cell* 71:663-70

13. Hu J, Anselmo D. 2000. In Vitro Reconstitution of a Functional Duck Hepatitis B Virus Reverse Transcriptase: Posttranslational Activation by Hsp90. *J.Virol.* 74:11447–55

14. Stahl M, Beck J, Nassal M. 2007. Chaperones Activate Hepadnavirus Reverse Transcriptase by Transiently Exposing a C-Proximal Region in the Terminal Protein Domain That Contributes to ε RNA Binding. *J.Virol.* 81:13354-64

15. Stahl M, Retzlaff M, Nassal M, Beck J. 2007. Chaperone activation of the hepadnaviral reverse transcriptase for template RNA binding is established by the Hsp70 and stimulated by the Hsp90 system. *Nucleic Acids Res* 35:6124-36

16. Lanford RE, Notvall L, Lee H, Beames B. 1997. Transcomplementation of nucleotide priming and reverse transcription between independently expressed TP and RT domains of the hepatitis B virus reverse transcriptase. *J.Virol.* 71:2996-3004

17. Weber M, Bronsema V, Bartos H, Bosserhoff A, Bartenschlager R, Schaller H. 1994. Hepadnavirus P Protein Utilizes a Tyrosine Residue in the TP Domain To Prime Reverse Transcription. *J.Virol.* 68:2994-9

18. Zoulim F, Seeger C. 1994. Reverse Transcription in Hepatitis B Viruses Is Primed by a Tyrosine Residue of the Polymerase. *J.Virol.* 68:6-13

19. Shin M, Lee J, Ryu W. 2004. A Novel cis-Acting Element Facilitates Minus-Strand DNA Synthesis during Reverse Transcription of the Hepatitis B Virus Genome. *J.Virol.* 78:6252–62

20. Ganem D, Varmus HE. 1987. The Molecular Biology of the Hepatitis B Virus. *Ann Rev Biochem* 56 651-93

21. Das K, Xiong X, Yang H, Westland CE, Gibbs CS, et al. 2001. Molecular Modeling and Biochemical Characterization Reveal the Mechanism of Hepatitis B Virus Polymerase Resistance to Lamivudine (3TC) and Emtricitabine (FTC). *J.Virol.* 75:4771-9

22. Hu J, Boyer M. 2006. Hepatitis B Virus Reverse Transcriptase and ε RNA Sequences Required for Specific Interaction In Vitro. *J.Virol.* 80:2141-50

23. Beck J, Nassal M. 1998. Formation of a Functional Hepatitis B Virus Replication Initiation Complex Involves a Major Structural Alteration in the RNA Template. *Mol.Cell.Biol.* 18 6265-72

24. Schaaf SG, Beck J, Nassal M. 1999. A Small 2'-OH- and Base-dependent Recognition Element Downstream of the Initiation Site in the RNA Encapsidation Signal Is Essential for Hepatitis B Virus Replication Initiation. *J.Biol.Chem.* 274:37787-94

25. Flodell S, Cromsigt JAMTC, Schleucher J, Kidd-Ljunggren K, Wijmenga SS. 2002. Structure Elucidation of the Hepatitis B Virus Encapsidation Signal by NMR on Selectively Labeled RNAs. *J.Biomol.Struct.Dyn.* 19:627-36

26. Ampt KAM, Werf RMvd, Nelissen FHT, Tessari M, Wijmenga SS. 2009. The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding. *Biochemistry* 48:10499-508

27. Flodell S, Petersen M, Girard FC, Zdunek J, Kidd-Ljunggren K, et al. 2006. Solution structure of the apical stem–loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res* 34:4449-57

28. Petzold K, Duchart E, Flodell S, Larsson G, Kidd-Ljunggren K, et al. 2007. Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res* 35:6854-61

29. Tompa P, Fuxreiter M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem.Sci.* 33:2-8

30. Leulliot N, Varani G. 2001. Current Topics in RNA-Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture. *Biochemistry* 40:7947-56

31. Lukavsky PJ, Kim I, Otto GA, Puglisi JD. 2003. Structure of HCV IRES domain II determined by NMR. *Nature Struct.Biol.* 10:1033-8

32. Lukavsky PJ, Puglisi JD. 2005. Structure Determination of Large Biological RNAs. *Methods Enzymol.* 394:399-416

33. Werf RMvd, Girard FC, Nelissen F, Tessari M, Wijmenga SS. 2008. ^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon. *Biomol NMR Assignments* 2:143-5

34. Ampt KAM, Werf RMvd, Nelissen FHT, Tessari M, Wijmenga SS. 2009. The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding. *Biochemistry* 48:10499-508

35. Wu B, Petersen M, Girard FC, Tessari M, Wijmenga SS. 2006. Prediction of molecular alignment of nucleic acids in aligned media. *J Biomol NMR* 35:103-15

36. Tucker BJ, Breaker RR. 2005. Riboswitches as versatile gene control elements. *Curr.Op.Struct.Biol.* 15:342-8

37. Boehr DD, Nussinov R, Wright PE. 2009. The role of dynamic conformational ensembles in biomolecular recognition. 5:789-96

38. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA, et al. 2008. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320:1471-5

39. Al-Hashimi HM, Walter NG. 2008. RNA dynamics: it is about time. *Curr.Op.Struct.Biol.* 18:321-9

40. Hermann T, Patel DJ. 2000. Adaptive Recognition by Nucleic Acid Aptamers. *Science* 287:820-5

41. Woessner DE. 1962. Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion. *J.Chem.Phys.* 37:647-54

42. Bailor MH, Sun XY, Al-Hashimi HM. 2010. Topology Links RNA Secondary Structure with Global Conformation, Dynamics, and Adaptation. *Science* 327:202-6

43. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450:1263-7

44. Zhang Q, Kim NK, Peterson RD, Wang ZH, Feigon J. 2010. Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc.Nat.Acad.Sci.USA* 107:18761-8
45. Popenda M, Szachniuk M, Blazewicz M, Wasik S, Burke EK, et al. RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *Bmc Bioinformatics* 11
46. Du Z, Ulyanov NB, Yu J, Andino R, James TL. 2004. NMR Structures of Loop B RNAs from the Stem-Loop IV Domain of the EnteroVirus Internal Ribosome Entry Site: A Single C to U Substitution Drastically Changes the Shape and Flexibility of RNA. *Biochemistry* 43:5757-71
47. Locker N, Easton LE, Lukavsky PJ. 2007. HCV and CSFVIRES domain II mediate eIF2 release during 80S ribosome assembly. *Embo J.* 26:795-805
48. Parsons J, Castaldi MP, Dutta S, Dibrov SM, Wyles DL, Hermann T. 2009. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* 5:823-5
49. Paulsen RB, Seth PP, Swayze EE, Griffey RH, Skalicky JJ, et al. 2010. Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc.Natl.Acad.Sci.USA*107:7263-8
50. Wang X, Qian X, Guo HC, Hu J. 2003. Heat shock protein 90-independent activation of truncated hepadnavirus reverse transcriptase. *J.Virol* 77:4471–80
51. Lin LJ, Wan F, Hu J. 2008. Functional and Structural Dynamics of Hepadnavirus Reverse Transcriptase during Protein-Primed Initiation of Reverse Transcription: Effects of Metal Ions. *J.Virol* 82:5703-14
52. Hu JM, Boyer M. 2006. Hepatitis B virus reverse transcriptase and epsilon RNA sequences required for specific interaction in vitro. *J.Virol* 80:2141-50
53. Hu JM, Flores D, Toft D, Wang XT, Nguyen D. 2004. Requirement of heat shock protein 90 for human hepatitis B virus reverse transcriptase function. *J.Virol* 78:13122-31
54. Haasnoot PCJ, Olsthoorn RCL, Bol JF. 2002. The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. *RNA* 8:110-22

4.7 Supplementary Material

Materials & Methods

Sample preparation

Samples of the 37nt. PBD RNA sequence were prepared as described earlier by Girard et al. (1). A sample with the RNA not isotope enriched and with RNA strand concentration of 0.19 mM was prepared in 90 % H_2O / 10 % D_2O in phosphate buffer (10 mM sodium phosphate pH 6.7, 0.1 mM EDTA). This sample will be referred as the non-labeled PBD H_2O sample. In addition, a similar sample was prepared in D_2O, referred to as the non-labeled PBD D_2O sample. A uniformly $^{13}C/^{15}N$ isotope-labeled sample was prepared. This sample has 1.0 mM RNA strand concentration, in H_2O (10% D_2O) in a phosphate buffer (10 mM sodium phosphate pH 6.7, 0.1 mM EDTA) to which sodium-2, 2-dimethyl-2-silapentane-5-sulfonate (DSS) was added (1.5 mM) as a chemical shift reference. The sample will be referred to as the labeled isotropic PBD sample. In addition, a second $^{13}C/^{15}N$ isotope-labeled sample was prepared as the one mentioned above, except that to this second sample Pf1 phages were added to a concentration of 17 mg/ml. It will be referred to as the labeled anisotropic PBD sample.

In a similar fashion, as described in (1) a fully $^{13}C/^{15}N$ isotope labeled samples of the 27 nt. ALD were prepared in a buffer of 93:7 H_2O:D_2O, 0,1 mM EDTA, 10mM sodium phosphate. The ALD-RNA sample had a concentration of 0.72 mM, and was made in duplo; one of the samples was dissolved in 15 mg/mL pf1 phages.

An unlabelled sample of 1.0 mM 62nt duck epsilon in 93 % H_2O / 7 % D_2O in phosphate buffer (10 mM sodium phosphate pH 6.7, 0.1 mM EDTA) and 1.5 mM DSS was produced in a similar fashion. Also a $^{13}C/^{15}N$ labeled sample of 0.67 mM 62nt duck epsilon in 93 % H_2O / 7 % D_2O in phosphate buffer (10 mM sodium phosphate pH 6.7, 0.1 mM EDTA) and 1.5 mM DSS was made. This sample was used to measure couplings. For RDC measurements, pf1 phages were added to a concentration of 8 mg/ml, and salts were added to preserve similar concentrations. The RNA concentration is this sample was 0.64 mM.

Samples of the human ε domains, PBH and ALH were prepared as described (1-3). Full-length human ε (62 nt) (FEH) was prepared in a similar fashion as FED.

NMR Spectroscopy Duck ε

Assignments PBD and ALD

Assignments of ^{1}H, ^{13}C and ^{15}N resonances of the primer loop of the epsilon element of the duck hepatitis B virus (PBD) and the apical loop of the epsilon element of the duck hepatitis B virus (ALD) were carried out as described by van der Werf et al. (4), and Ampt et al. (5), respectively. Assignments can be found in the BMRB under entry numbers 15786 (PBD), 15656 (ALD). These assignments were used directly to assign the spectra of the 62nt duck epsilon. Measuring imino-NOESYs with different mixing times validated these assignments:

75, 100 and 125 ms at different 15 °C and different mixing times. To obtain resonance assignments for FED several 2D NMR spectra were recorded of various FED samples as listed in Table S4.1; the procedure is briefly described in the main text. The assignments are given in Table S4.2.

RDCs of the PBD and ALD structure elements

Residual Dipolar Couplings (RDCs) were measured from the labeled anisotropic and isotropic PBD samples. Spectra for both RDC and relaxation measurements were recorded on a 600MHz Varian UNITY INOVA NMR spectrometer.

To measure RDCs, several region-specific IPAP and coupled 2D HSQC spectra were recorded on both the isotropic and anisotropic samples of PBD, i.e. at 25 °C coupled (^{13}C-^{1}H) CT-HSQCs for the region C1'H1', at 15 °C for the regions C2H2, C5H5 IPAP for pyrimidines, C6H6 IPAP for pyrimidines, a (^{15}N-^{1}H)-CTHSQC at 15°C to obtain N1H1(G) and N3H3(U) couplings. Furthermore, a 3D HCCH-TOCSY at 25 °C was recorded for measuring C1'H1',C2'H2', C3'H3', and C4'H4' couplings (6).

This set of experiments was repeated on the ALD-samples at 15 °C, to obtain RDCs for the ALD.

RDC measurements on the full 62 nt. Epsilon, FED

In a similar fashion as described above, RDCs were measured on the full 62 nt. Epsilon (FED) in 8mg/ml pf1 phages for C6-H6, C8-H8, C2-H2 and N1/3-H1/3. Coupled HSQC spectra of the isotropic sample and the anisotropic sample were measured on an 800MHz Varian Unity INOVA spectrometer at 25 °C. Since no tertiary or quaternary interactions of epsilon did occur, the assignments of the ALD and PBD could be used.

NMR restraints – PBD, ALD and FED (Table S4.3)

From NOESY spectra, distance restraints were obtained by relating the reference distance (H5/H6-correlation of pyrimidines) to a distance of 2.45 Å. The distances measured in PBD are intra residual contacts H1'-H2'/2/5/8/6, H2'-H2/5/8/6, inter residual contacts H1'$_{i}$-H2/6/8$_{i+1}$, H2'$_{i}$-H2/6/8$_{i+1}$ H5$_{i}$-H8$_{i+1}$. The correlations with H1'/H5 were observed in a NOESY with mixing time 500ms, and the correlations with H2' were observed in a NOESY with mixing time 30ms (25C, 800MHz Varian UNITY INOVA spectrometer, spectral width 12 kHz for both spectra). For the UUCG tetra-loop in the PBD, the NOE restraints and dihedral restraints were measured from the published structure of the UUCG-tetra loop (PDB entry 1HLX,) and checked against the experimental data. The sugar-puckering modes were derived qualitatively from the DQF-COSY (1) (25C, 800MHz Varian UNITY INOVA spectrometer, spectral width 12kHz).

In total, 28 N-H and 60 C-H RDCs were collected, from the PBD sample dissolved in 17 mg/mL Pf1-phages. For the apical loop, 98 RDCs were collected: 9 N-H and 89 C-H couplings from a sample dissolved in 15 mg/mL pf1-phages.

Backbone dihedral restraints within the helices were defined as described in Gelbin et al. (7), the α, and ζ angles are defined as –sc, and γ as +sc with a range of 20 degrees flexibility. For the UUCG-loop (residues 20-24) values were obtained from the structure by Allain et al (8) (PDB-entry: 1HLX) and had a 30 degrees range of flexibility, and the primer loop region (residues 7-13) was left unrestrained in terms of backbone dihedral angles.

The sugar-puckering dihedral restraints, as derived qualitatively from the DQF-COSY, restrained all residues to an N-puckering, the following residues are S-puckered: 9-11, 21-22,30 in the PBD, and residues 14-16 in the ALD, and for residue 25 in the ALD no pucker mode could be derived. Dihedrals are chosen as defined by Gelbin et al. (7)

For the UUCG sugar-dihedrals were obtained from the structure by Allain et al (8) (PDB-entry: 1HLX).

The flexibility of the sugars had a range of 5 degrees, except for the primer loop region and the UUCG loop region in the PBD where the restraints are somewhat looser.

All χ angles, except residue 23 which is a *syn*-conformer, are restrained to anti-conformers values are taken from the paper by Gelbin (7).

An overview of the number of derived NMR restraints is provided in Table S4.3. NOE, RDC, and dihedral restraints were submitted to the BMRB and can be found under accession code: 15786 (PBD), 15656 (ALD).

Human ε

Assignment for ALH and PBH were obtained as described (1; 3; 9); for FEH the assignments were obtained in a way similar to that used in FED (Fodell, Schleucher, Petersen, Wijmenga, personal communication).

Structure analysis

Structure analysis of HISXSYHII RNAs (Bailor et al., Science 2010)

1. Reference frame definition to calculate Euler angles between helix I and II

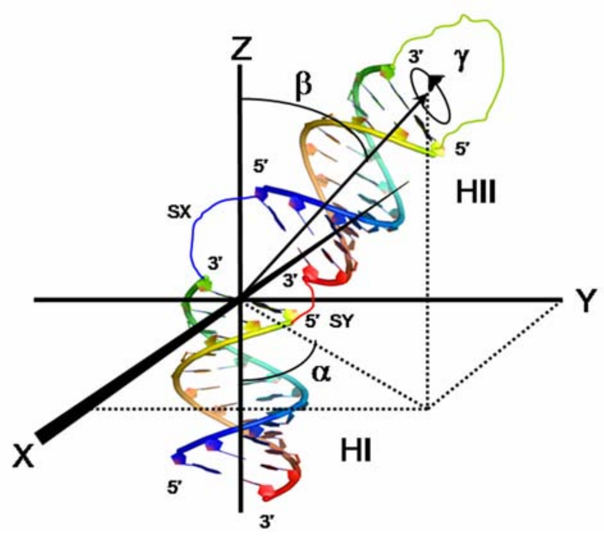


Figure S4.1 Reference axis frame for two-way junction HISXSYHII. Definitions for two-way junctions are given as Bailor et al.(10) IUPAC definitions on junction terminology can be found in Lilley et al.(11; 12). Moreover, junctions and terminology in RNA and DNA have been reviewed by Lilley (13; 14).

The RNA molecules considered are two-way junctions (H2; (11; 12)) and designated as HISXSYHII (Figure S4.1). Here, HI and HII are the lower and upper helix. SX refers to the single-stranded region of X nucleotides that connects the 5'-strand of HI with the 5'-strand of HII. SY refers to single-stranded region of Y nucleotides in the opposing strand. X is taken to be larger than Y. The Euler angles α, β, and γ between the helices as used by Bailor et al. (10; 15; 16) are defined in Figure S4.1. They are defined following the recent IUPAC conventions as given by Dickerson et al. (17). Specifically, for HI and HII each, a global helix axis is defined (18-20). Of the polar angles β and α, the β-angle is called inter-helix angle, and α the helical twist angle of HI. The γ-angle designates rotation around the helix axis of HII, i.e. helical twist of HII. The total helical twist equals α+γ. The reference axis frame XYZ in Figure S4.1 is attached to helix HI and defined as follows. The Z-axis lies parallel to the global helix axis of HI (18-20) and points in the direction to which 5'-strand travels. Before defining the X- and Y-axis, we first discuss the local axis frame, X', Y', and Z' defined by a basepair (21).

The right-handed local basepair coordinate frame is attached to a base and drawn following the qualitative guidelines by Dickerson et al. (17) The X’-axis is perpendicular to the bisector of the C1’-C1’ vector spanning the basepair, and points in the direction of the major groove. (18; 21) The Y’-axis runs along the long axis of the idealized base-pair in the direction of the sequence strand, parallel with the C1’-C1’ vector, and is displaced so as to pass through the intersection of the X’-axis and the line connecting the pyrimidines Y(C6) and purine R(C8) atoms (21). The Z’-axis is defined by the right-handed rule, i.e. Z’ = X’ x Y’. For right-handed A and B-DNA, the Z’-axis accordingly points along the 5’ to 3’ direction of the sequence strand (21).

The global helix axis, the Z-axis, is obtained from a least squares fit (22) through all local helix axes, the Z’-axes (20). The X- and Y-axis can conveniently be defined based on the local helix axis system of a chosen basepair. The X-axis is defined by the right-handed rule from X = Y’ x Z and the Y-axis from Y =Z x X. We chose for the definition of the X- and Y-axis of the global helix frame, the local basepair frame of the HI basepair closest to the 5’-end of the SX strand.

Structure calculation

Primer binding loop, PBD: structure calculation (Table S4.7)

For structure calculation the following protocol has been designed; it initially neglects potential dynamics, but in a later takes dynamics into accounts.

I. From NOEs and chemical shifts, the extent of the helical elements was identified. The PBD can be classified as a two-way junction, HISxSyHII. Here HI is lower helix and HII is the upper helix, which is capped by a UUCG tetraloop. The helices HI and HII are connected via the primer loop.

II. The confirmed helical elements were built in 3DNA (23) based on fiber diffraction model A-helix parameters (23, and references therein), from here on referred to as standard helices. The correctness of standard A-helix conformation was checked by back calculation with NUCHEMICS of the ^{1}H chemical shifts (24) and subsequent comparison with the experimental ^{1}H chemical shifts. In addition, it was checked whether the helix elements violated any of the observed NOE restraints. No violations were observed.

III. a) The newly built helical elements were then combined with 1) a UUCG loop from the online protein databank (PDB entry code: 1HLX) as published earlier by Alain and Varani (8; 25) and 2) the primer loop; the primer loop was generated in Xplor v3.851 (26) as an extended structure. b) The complete structure was obtained via a torsion angle dynamics protocol as published by Wu et al. (27), in which the helix stems, the primer loop and the UUCG tetra loop acted as input. During this dynamics run, the aromatic planes of the helices were kept fixed; the UUCG-loop was treated as a rigid body, by means of artificial NOE distance restraints, while the primer loop was allowed to change conformation. In this manner, 300 starting structures were calculated. These starting structures have the correct helix conformations and a correct UUCG tetraloop conformation; the primer loop is randomized and at the same time this implies that the relative orientation of the two helices is randomized.

The energy of each starting structure was then calculated and it was checked whether any displayed NOE and holonomic constraint violations. The structures with no NOE and holonomic restraints violations were selected and the one with the lowest overall energy was submitted to IV.

We note in passing that we cannot reliably check the starting structures against the RDCs, because the relative helix orientations are not known at this stage and thus the alignment tensor parameters. We note further with regard to the torsion angle dynamics (TAD) protocol employed the following. During this TAD protocol, the helices were kept fixed, and 7603 artificial distance restraints kept the UUCG loop in its correct conformation, to solve interface problems between the helical stem and the UUCG loop. Note that in the TAD protocol the C3’-C4’ bonds are cut, while the base planes are kept fixed. In the TAD the structure is heated from 300 to 20,000 K in 102 steps of 0.015 ps, during these stages the NOEs (force constant 50 kcal $Å^{-2}$ mol^{-1}) and dihedrals are employed as restraints (force constant 5 150 kcal $(°)^{-2}$ mol^{-1}). Subsequently, the structure is cooled to 1,000 K in 474 steps of 0.015 ps, the NOE force constant is increased to 150 kcal $Å^{-2}$ mol^{-1}. Then, the structure is cooled further to 300K using verlet dynamics for 515 steps of 0.0005 ps, while increasing the dihedral force constant to 200 kcal $(°)^{-2}$ mol^{-1}. The Verlet dynamics continues at 300K for during 15000 steps of 0.0005 ps with the NOE force constant set to 50 kcal $Å^{-2}$ mol^{-1}. Finally, a Powell minimization is applied for 10,000 steps, keeping the parameters as set in the last Verlet dynamics part.

IV. Helix reorientation as derived from RDCs. The relative helix orientation of the two helix elements (residues 1-6, 32-37, and 14-29) were derived from RDCs using a grid search. Seventy RDCs were used, i.e. all RDCs in the helical stems and in the UUCG tetra loop, excluding the RDCs of the 2nd U and C in the tetra loop, as well as the RDCs of the primer loop. The lower helix stem was kept fixed (both orientation and conformation). The upper helix and tetra-loop was rotated as a rigid body, around its three Euler angles to find the optimum orientation. To eliminate the primer loop, the strand was cut between residues 6 and 7 and between residues 11 and 12, so that primer loop is excised and does not affect RDC calculations. The Phosphate between residue 30 and 31 was used as pivot point. During the grid search, the RDCs are calculated from the intermediate structures based on their gyration tensor derived from the phosphate groups so that the alignment tensor conforms to the structure at all times (28). This eliminates five adjustable parameters (29). The optimum structure is defined as the one that gives minimum squared difference between the measured and calculated RDCs. This structure with the correct relative helix orientations and RDC calculation also provides the correct alignment tensors axial component and rhombicity as well as its orientation.

V. From the structure with correct relative helix orientation, a family of 267 starting structures for RDC refinement was generated using a torsion angle dynamics protocol as described earlier, except that now the helix orientations are kept fixed; only the primer loop needs to be incorporated and refinement of the conformation of residue 30 (pivot point) is required.

VI. The resulting 267 structures were subjected to an RDC refinement, in which 5 refined structures were calculated from each of the 267 starting structures.

In the helices (residues 1-8, 14-19, 24-29, 31-37) the bases were kept fixed. Only intra- and inter-residual NOE distance restraints were used in the primer loop region, because no long-range NOE restraints were found from within this region. Note again that RDCs of mobile residues were excluded (based as described above on peak intensities). In this way, the primer loop structure is based on RDCs, sequential and intra-residual distance restraints and context provided by the overall structure which the correct overall relative helix orientation.

The details of the RDC refinement protocols are as follows. During the protocol the axial component and rhombicity of the alignment tensor were kept fixed to the ultimate optimal values derived under IV) (see Wu et al. 2006); the alignment tensor orientation was left floating. The RDC-refinement protocol takes place in a temperature bath of 300K. It starts out with 2.0 ps (4000 steps) of Verlet dynamics period with a force constant of 20 kcal $(°)^{-2}$ mol^{-1} for dihedrals, and 0.2 kcal Hz^{-2} mol^{-1} for RDCs. The force constant for NOE restraints is 5 kcal $Å^{-2}$ mol^{-1}. Subsequently, the Verlet dynamics is continued during which in 50 cycles of 0.5 ps (1000 steps each) the RDC, NOE and dihedral force constants are increased to 0.4 kcal Hz^{-2} mol^{-1}, 20 kcal $Å^{-2}$ mol^{-1}, and 200 kcal $(°)^{-2}$ mol^{-1}, respectively. The Verlet dynamics is then continued for another 20.0 ps (20,000 steps) with the RDC force constant set back to 0.2 kcal Hz^{-2} mol^{-1}, while the dihedral and NOE force constant are kept at the same values. The script finishes with 1000 steps Powell minimization.

An overview was made of resulting structures that did not show any NOE violations (>0.5 Å), RDC violation (>10 Hz), or dihedral angle violation (>5 degrees). From these structures the rhombicity of the alignment tensor were calculated via the gyration tensor method (28). Ideally, this would result for each separate structure in a rhombicity of 0.10, the value that was used in the foregoing calculation.

Further more the chemical shifts of the calculated structures were calculated from the structures using NUCHEMICS (24) and the RMSD with respect to measured values was calculated for helical parts (residues 2-7,15-20,23-28,31-36), and for the primer loop (residues 9-13 and 30).

From the overview of structures that did not show any violations, their total energy, their RDC-energy resulting from structure calculation, together with their chemical shift RMSD for helical parts, and for bulge parts, criteria were set up to select the structures best matching our experimental NMR data. These criteria are: overall energy < 1200 kcal/mole, RDC energy < 220 kcal/Hz, Rhombicity < 0.15, $RMSD(\delta_{helix})<0.172$ ppm, and $RMSD(\delta_{primer\ loop})<0.370$ ppm.

The 7 structures matching all these criteria were validated using NUCHECK(7; 30; 31), and did not show any corruptions. The ensemble of 7 structures will be submitted to the PDB.

Apical stem loop, ALD: structure calculation

The structure calculation protocol used is essentially as described by Ampt et al. (32). Briefly, structure calculations were carried out with X-PLOR 3.851 using a torsion angle dynamics protocol (1). Starting from an extended structure, 500 structures were generated using classical NMR restraints (Table 4.1). The twenty lowest energy structures were submitted to a second round of torsion angle dynamics, generating 20 refined structures out of each selected structure. From the resulting 400 refined structures, the 20 with lowest energy were selected for RDC refinement, all with NOE violations < 0.5 Å and with dihedral violations < 5°. For each of these starting structures the rhombicity and the axial component of the alignment tensor were calculated using the method proposed by Wu et al. (28) as well as with the software Module (28; 33). Values of Da = - 25 Hz and R = 0.08 were used to calculate 5 refined structures from each starting structure resulting in100 RDC refined structures. The ten lowest structures using classical restraints have been submitted to the RCSB protein data bank as 2K5Y. The ten lowest energy structures after RDC refinement have been submitted as 2K5Z.

Full-length 62nt duck HBV ε, FED: structure calculation (Table S4.7)

PBD and ALD were aligned on their shared basepairs: PBD 14-29, and ALD 3-27. The first two basepairs in the ALD are removed before assembling the complete 62nt, FED structure. Besides this, the bottom helix of the 62 nt full epsilon, containing the 3’- and 5’-end of the RNA strand, was built in 3DNA (23) based on a fiber model. The top U-G and A-U basepair of this helix was aligned with the one of the PBD. Subsequently the full 62 nt duck epsilon was built by replacing the top helix of PBD by the structure of the ALD, and the bottom helix is replaced by the helix generated in 3DNA.

Helix orientations were obtained from RDCs. For refinement of RDCs a series of orientations of both helical elements within the structure were generated, by rotating the apical stem over Euler angles (α,β,γ) with residue 49 P acting as a pivot point and minimizing the RDC RMSD within a certain range. The Euler angles are described in the paper by Zhang et al. (34) and have a range of -90° to +40° for the α angle, -60° to 0° for the β angle, and -60° to +90° for the γ angle. From relaxation data, it was determined that helix domain motion was present to a degree S^2 is circa 0.8. This domain motion can be modeled as motion in a cone with a half angle of circa 20°, i.e. the upper helix undergoes motion in a cone with respect to the bottom helix. This motion can be modeled simply by rotation over the Euler angles and while tracking the RMSD of the RDCs. The $RMSD_{RDC}$ was set to circa 10 Hz, which is close to $3\sigma_{RDC}$ with σ_{RDC} the experimental RDC error. This setting leads to a cone half-angle of circa 20°, after the subsequent refinement.

The subsequent structure refinement protocol started from these assembled 62nt FED models. In this subsequent structure calculation protocol, the following RDCs were used: RDCs from the primer loop (23 RDCs), RDCs from the apical loop (80 RDCs), and RDCs measured on the 62 nt epsilon in phages (36 RDCs). The protocol uses a different alignment tensor for each of the three sets of RDCs, implemented in the protocol as floating alignment tensors. For each starting structure in the refinement the axial component and rhombicity for the RDC set of the 62nt epsilon in phages were calculated using the method proposed by Wu et al. (28). The other axial components are kept the same as in foregoing calculations. Further the calculation protocol is exactly as described under ‘PBD’ as RDC refinement protocol, while keeping the aromatic planes of following residues of the bottom helix fixed: 1-2, 5-6, 9-11, 13, 50, 52-54, 57-58, 61-62. This results in a set of model structures, which satisfies holonomic restraints as validated by NUCHECK (31). The resulting structure was validated using NUCHECK. The ensemble of 68 structures will be submitted to the PDB.

Human epsilon

The structure calculation for the human ε apical stem loop was described in Flodell et al. (2); the structure of the primer binding loop of human ε, PBH, was carried out in a similar fashion, while the structure calculation of full-length 62nt human ε, FEH, followed a strategy largely similar to the one employed for FED.

NMR relaxation measurements - dynamics

Duck epsilon

^{13}C relaxation measurements on PBD (Table S4.4)

^{13}C Relaxation measurements, R_1, $R_{1\rho}$, and steady-state NOE, were carried out on a separately prepared 1.0 mM $^{13}C/^{15}N$ enriched sample (90% H_2O, 10% D_2O, 10 mM sodium phosphate pH 6.7, 0.1 mM EDTA), referred to as the labeled PBD relaxation sample. To measure relaxation rates spectra were recorded on the labeled PBD relaxation sample at 25 °C, and on the labeled ALD sample at 15 °C. For R_1-measurements HSQCs were recorded, in which magnetization C_z is stored along the z-axis for a time T_a while decoupling to prevent proton cross-relaxation. For measurement of $R_{1\rho}$ the magnetization was stored as $C_{x/y}$ under a spin lock field of ~4.4 kHz for C2 and C8, ~1.3 kHz for C1' and 2.7 kHz for C5, during a time T_b. For measurement of steady state NOE a presaturation was applied during 40 ms, and a reference was recorded without presaturation. The experiments are modified HSQC experiments employing pulsed field gradients (35-37).

Spectra were recorded with T_a values (20,100,200,300,400,500,600,800) ms, T_b values (4, 8, 12, 16, 20, 24, 28, 32) ms. All measurements were recorded in duple. The spectra fitted with a nonlinear least square fit using NMRPipe (38), from resulting data R_1 and $R_{1\rho}$ values were collected, and $R_{1\rho}$ were corrected for off-resonance effects. Relaxation rates were recorded for the carbon nuclei: C1', C2, C5 and C8. The relaxation rates are listed in Table S4.4.

All spectra were processed in NMRPipe and the chemical shift was calibrated with respect to DSS.

Relaxation measurements on ALD

The relaxation measurements on ALD were carried as described in Ampt et al. (32). Note that for ALD on the H2 protons of the ALD relaxation dispersion measurements were recorded at a Varian UNITY Inova spectrometer operating at 600MHz. The sample was at 35 °C and 37.5 °C, and $R_{1\rho}$ were recorded using a standard constant-time cross relaxation compensated CPMG scheme (39; 40). The constant time was set to 20ms, and the CPMG field strength was increased from 100Hz to 1400Hz. Experiments were recorded in a 1D fashion.

Analysis of relaxation data and dynamics of PBD and ALD

Primer bulge, PBD (Table S4.5)

The relaxation data were analyzed using MODELFREE v4.20. Relaxation rates had standard errors of 2.0%, 4.0%, and 5.0% for R_1, $R_{1\rho}$, and NOE respectively. For the C-H bond motional averaged bond lengths were taken:1.115 Å for C1'H1', and 1.104 Å for aromatic C-H vectors (10), and the chemical shift anisotropies as described in Stueber et al. (41) C2 150.0 ppm, C8 134.0 ppm, C5 138.0 ppm, and C1' 45.0 ppm as described by Duchart et al. (42).

An initial guess of the fully axial symmetric diffusion tensor was made in HYDRONMR (43) from the ten best structures from the calculation, the result was optimized with MODELFREE (44; 45) by grid search followed by a Powell minimization.

The rotational correlation time was found to be 9.76 ns, and the anisotropy of the diffusion tensor 1.9, this matches with HYDRONMR data. Model selection was done as described in the paper by Mandel et al. (44) using only model 1, 2, and 5. Model 1 only fits S^2, model 2 both S^2 and τ_e, and model 5 S_s^2,S_f^2, and τ_e in the (extended) Lipari-Szabo approximation (46). If these models did not fit, we used model 4, which fits S^2, τ_e and R_{ex}. Results of our analysis are reported in Table S4.5.

Apical loop, ALD

The analysis was carried out as describe din Ampt et al. (32). Briefly, the rotational correlation time was found to be 9.4 ns from MODELFREE analysis, and the anisotropy of the diffusion tensor 2.7; values that match with the HYDRONMR data. Model selection was done as described in the paper by Mandel et al. (44) using only model 1, 2, and 5. Model 1 only fits S^2, model 2 both S^2 and τ_e, and model 5 S_s^2,S_f^2, and τ_e in the (extended) Lipari-Szabo approximation (46). If these models did not fit, we used model 4, which fits S_2, τ_e and R_{ex}. For some models model 3 S^2, R_{ex} was found, however no exchange was found when using this model, a more complicated model was chosen according to the flowchart in the paper by Mandel et al. (44).

Full-length duck HBV ε, FED: back-calculation of S^2 from the ensemble of structures (Table S4.6).

First, the 68 structures calculated for the 62 nt. epsilon of the duck hepatitis B virus were aligned. Next, the generalized order parameters of the full 62nt epsilon were calculated using the method described by Best et al. (47), which in essence resembles the method by Lipari et al. (48). In these calculations, we assumed that our ensemble represents a complete set, covering the entire helix motion. So the generalized order parameter can be approximated from the auto-correlation function in a modelfree approach via:

$$S^2 = \lim_{t\to\infty} C_i(\tau) = \lim_{t\to\infty} \left\langle P_2\left(\mu(0)\bullet\mu(\tau)\right)\right\rangle \tag{4.1}$$

Where μ is the unit vector along the bond considered in the molecular coordinate frame(48). Now, by defining a matrix ϕ (49), the order parameter can be calculated from equation (4.3).

$$\phi_{ij} = \frac{\vec{r}_i \bullet \vec{r}_j}{\left|\vec{r}_i\right| \cdot \left|\vec{r}_j\right|} \quad \vee \quad \{i,j\} \in \{x,y,z\} \tag{4.2}$$

$$S^2 = \frac{3}{2} \cdot tr\left\langle \phi \right\rangle^2 - \frac{1}{2} \cdot \left(tr\left\langle \phi \right\rangle \right)^2 \tag{4.3}$$

Here r_i and r_j are vectors to the heteronucleus and its attached proton in the pdb-frame. Matrix ϕ basically contains the dot products of the bond vectors concerned. In our case this concerns the following C-H bonds: C1’H1’, C2H2, C5H5, C6H6, C8H8. Results are shown in Table S4.6.

To relate experimental S^2 determination following from relaxation measurements to that of the 62nt duck epsilon, the found order parameters in ALD (32) are multiplied by $<S^2_{\text{helix motion}}>=0.88$ to account for the motion of the two separate stems in the structure. This value was found by minimizing the average found generalized order parameter of the apical stem to the generalized order parameter of the bottom stem.

Table S4.1. Spectra recorded on 62 nt FED

T (°C)	**Conditions**	**Fields (MHz)**	**Experiment**	**Notes**
5	unlabeled, H_2O	800	imino-NOESY	mixing times 150, 200ms
15	unlabeled, H_2O	800	imino-NOESY	mixing times 75, 100, 125, 150 ,200, 300, and 500ms
15	uniform ^{13}C, ^{15}N labeled, H_2O	600	[^1H,^{15}N]-HMQC	decoupled
15	uniform ^{13}C, ^{15}N labeled, H_2O	500,600,800	[^1H,^{15}N]-HSQC	coupled
15	uniform ^{13}C, ^{15}N labeled, H_2O,Pf1(8mg/ml)	800	[^1H,^{15}N]-HSQC	coupled
25	uniform ^{13}C, ^{15}N labeled, H_2O	600	[^1H,^{13}C]-CTHSQC	decoupled, regions: C2, C5 ,C6/8
25	uniform ^{13}C, ^{15}N labeled, H_2O	600,800	[^1H,^{13}C]-CTHSQC	coupled, regions: C2, C5, C6/8
25	uniform ^{13}C, ^{15}N labeled, H_2O,Pf1(8mg/ml)	800	[^1H,^{13}C]-CTHSQC	coupled, regions: C2, C5, C6/8

NMR spectra of FED and the conditions under which they were recorded

Table S4.2. NMR Assignments of FED (in ppm)

	5 °C			15 °C					25 °C					
	H1	H2	H3	H1	H2	H3	N1	N3	H2	H6	H8	C2	C6	C8
G1	-	-	-	12.780	-	-	-	-	-	-	-	-	-	-
G2	13.310	-	-	13.190	-	-	148.0	-	-	-	-	-	-	-
U3	-	-	14.000	-	-	13.950	-	162.7	-	-	-	-	-	-
G4	10.800	-	-	10.780	-	-	144.2	-	-	-	-	-	-	-
A7	-	6.525	-	-	6.561	-	-	-	6.573	-	-	152.0	-	-
A8	-	6.995	-	-	7.023	-	-	-	7.031	-	-	152.6	-	-
A9	-	7.366	-	-	7.396	-	-	-	7.407	-	-	153.2	-	-
G10	12.880	-	-	12.870	-	-	147.3	-	-	-	6.972	-	-	135.6
G11	13.210	-	-	13.210	-	-	148.3	-	-	-	7.098	-	-	136.1
U12	-	-	13.600	-	-	13.590	-	161.8						
U14*	-	-	-	-	-	-	-	-	-	7.503	-	-	140.9	-
U16*	-	-	-	-	-	-	-	-	-	7.660	-	-	141.8	-
A19*	-	-	-	-	-	-	-	-	7.981	-	8.229	155.1	-	141.4
G21*	13.150	-	-	13.100	-	-	147.9	-	-	-	-	-	-	-
U22*&	-	-	14.370	-	-	14.340	-	162.6	-	7.767	-	-	141.5	-
U24&	-	-	-	-	-	11.150	-	157.8	-	-	-	-	-	-
A25&	-	7.450	-	-	7.488	-	-	-	7.513	-	8.261	153.9	-	140.5
A27&	-	-	-	-	7.668	-	-	-	7.410	-	7.883	153.3	-	139.4
U28&	-	-	11.870	-	-	11.820	-	158.1	-	-	-	-	-	-
U29&	-	-	11.620	-	-	11.620	-	158.7	-	7.950	-	-	143.1	-
G30&	10.840	-	-	10.820	-	-	144.1	-	-	-	7.698	-	-	137.1
U32&	-	-	-	-	-	10.880	-	157.7	-	-	-	-	-	-
G33&	-	-	-	-	-	-	-	-	-	-	7.729	-	-	140.8
U34&	-	-	-	-	-	-	-	-	-	7.673	-	-	144.2	-
U35&	-	-	-	-	-	-	-	-	-	7.822	-	-	144.2	-
G36&	-	-	-	13.210	-	-	148.3	-	-	-	-	-	-	-
G39&	11.560	-	-	11.510	-	-	144.8	-	-	-	-	-	-	-
U40&	-	-	13.830	-	-	13.770	-	-	-	-	-	-	-	-
G41&	12.520	-	-	12.530	-	-	147.3	-	-	-	7.668	-	-	136.4
U42&	-	-	13.700	-	-	13.660	-	161.7	-	-	-	-	-	-
G43&	-	-	-	-	-	-	-	-	-	-	7.780	-	-	138.1
G45&	12.550	-	-	12.530	-	-	-	-	-	-	7.747	-	-	137.4

A46*&	-	7.825	-	-	7.850	-	-	-	7.853	-	7.941	154.0	-	139.9
C47*	-	-	-	-	-	-	-	-	-	7.119	-	-	136.1	-
G49*	-	-	-	-	-	-	-	-	-	-	7.909	-	-	138.4
U50	-	-	-	-	-	14.110	-	-	-	-	-	-	-	-
A51	-	7.188	-	-	7.214	-	-	-	7.221	-	-	153.5	-	-
U54	-	-	13.880	-	-	13.870	-	162.1	-	-	-	-	-	-
U55	-	-	13.830	-	-	13.800	-	162.6	-	-	-	-	-	-
U56	-	-	13.080	-	-	13.060	-	162.1	-	-	-	-	-	-
G57	11.950	-	-	11.960	-	-	146.4	-	-	-	-	-	-	-
G58	13.240	-	-	13.250	-	-	148.3	-	-	-	7.165	-	-	135.8
U59	-	-	11.690	-	-	11.700	-	158.7	-	-	-	-	-	-
A60	-	7.284	-	-	7.318	-	-	-	7.331	-	-	153.4	-	-

* Residue whose resonance positions overlap with PBD

& Residues whose resonance positions overlap with ALD

Residues that are not mentioned in the list were not assigned. Resonance positions of FED that significantly differ from the resonance position of the corresponding resonance in PBD are marked in bold, and resonance positions of FED that significantly differ from the resonance position of the corresponding resonance in ALD are marked in italic. Differences in exact resonance positions were observed between PBD and FED in the beginning of the apical stem. In the bottom of the apical stem in FED, residues U22 and A46 do not completely overlap with the ALD resonances U3 and A27. These differences are most likely an effect due to the change of the top-stem into a true apical stem. For PBD the top loop has changed, and for ALD the bottom of the stem is connected to a primer bulge.

Table S4.3. Number of NMR restraints for ALD, PBD, and full 62 nt. epsilon

	ALD	PBD	62 nt. epsilon
NOE			
H-bonds	59	0	119
intranucleotide	125	210	120
internucleotide	93	149	149
total	277	359	388
RDC			
D_a	-25	-46	[-19.64,-16.36]
R	0.08	0.1	[0.06,0.34]
C-H	89	60	111
N-H	9	28	28
total	98	88	139
Dihedrals			
Backbone	158	146	269
Sugar	140	311	548
Chi	28	37	58
total	326	494	875
planarity’s	12	0	12
grand total	713	941	1414

Table S4.4. Relaxation data as observed from 37nt PBD

	C1’			C5(pyr)/C8(pur)			C2		
Residue	R_1	$R_{1\rho}$	NOE	R_1	$R_{1\rho}$	NOE	R_1	$R_{1\rho}$	NOE
1									
2				1.71 ± 0.03	37.9 ± 1.5	1.22 ± 0.06			
3	1.27± 0.03	33.8 ± 1.4	1.25±0.06	1.89 ± 0.04	35.9 ± 1.4	1.19 ± 0.06	1.90 ± 0.04	41.4 ± 1.7	1.24 ± 0.06
4	1.34± 0.03	31.0 ± 1.2	1.35±0.07	2.07 ± 0.04	38.6 ± 1.5	1.17 ± 0.06			
5				1.85 ± 0.04	35.0 ± 1.4	1.20 ± 0.06			
6									
7	1.62± 0.03	32.6 ± 1.3	1.36±0.07	2.05 ± 0.04	34.3 ± 1.4	1.30 ± 0.07			
8				2.14 ± 0.04	35.7 ± 1.4	1.11 ± 0.06			
9	2.04± 0.04	29.4 ± 1.2	1.4±0.07						
10				2.52 ± 0.05	20.3 ± 0.8	1.50 ± 0.08			
11	2.20± 0.04	27.5 ± 1.1	1.46±0.07	2.35 ± 0.05	24.4 ± 1.0	1.41 ± 0.07			
12				1.88 ± 0.04	29.7 ± 1.2	1.31 ± 0.07	2.00 ± 0.04	36.7 ± 1.5	1.27 ± 0.06
13				2.13 ± 0.04	29.6 ± 1.2	1.31 ± 0.07			
14				1.74 ± 0.04	37.7 ± 1.5	1.22 ± 0.06			
15				1.83 ± 0.04	42.6 ± 1.7	1.18 ± 0.06			
16	1.70± 0.03	29.4 ± 1.2	1.43±0.07	1.89 ± 0.04	40.4 ± 1.6	1.11 ± 0.06			
17				2.04 ± 0.04	38.0 ± 1.5	1.22 ± 0.06			
18	1.39± 0.03	33.2 ± 1.3	1.28±0.06	1.76 ± 0.04	36.5 ± 1.5	1.20 ± 0.06			
19	1.32± 0.03	28.3 ± 1.1	1.39±0.07						
20	1.40± 0.03	25.5 ± 1.0	1.26±0.06	2.27 ± 0.04	25.5 ± 1.0	1.26 ± 0.06			
21	1.31± 0.03	30.3 ± 1.2	1.34±0.07	1.74 ± 0.04	27.9 ± 1.1	1.29 ± 0.06			
22	1.21± 0.02	34.4 ± 1.4	1.37±0.07	1.83 ± 0.04	41.9 ± 1.7	1.22 ± 0.06			
23	1.27± 0.03	26.8 ± 1.1	1.31±0.07	1.63 ± 0.03	35.6 ± 1.4	1.25 ± 0.06			
24				1.76 ± 0.04	34.8 ± 1.4	1.27 ± 0.06			
25	1.40± 0.03	28.8 ± 1.2	1.25±0.06	1.93 ± 0.04	39.1 ± 1.6	1.19 ± 0.06			
26									
27				1.68 ± 0.03	39.2 ± 1.6	1.22 ± 0.06			
28	1.43± 0.03	38.6 ± 1.5	1.23±0.06	1.60 ± 0.03	38.6 ± 1.5	1.20 ± 0.06			
29	1.4 ± 0.03	27.8 ± 1.1	1.28±0.06	1.91 ± 0.04	40.6 ± 1.6	1.24 ± 0.06			
30				2.06 ± 0.04	32.4 ± 1.3	1.31 ± 0.07			

31	1.53 ±0.03	34.6 ± 1.4	1.32±0.07	1.72 ± 0.03	37.0 ± 1.5	1.22 ± 0.06
32				1.93 ± 0.04	39.9 ± 1.6	1.20 ± 0.06
33				1.90 ± 0.04	41.6 ± 1.7	1.14 ± 0.06
34				1.86 ± 0.04	38.5 ± 1.5	1.27 ± 0.06
35				2.03 ± 0.04	39.0 ± 1.6	1.23 ± 0.06
36				2.27 ± 0.05	31.4 ± 1.3	1.37 ± 0.07
37				2.02 ± 0.04	36.1 ± 1.4	1.40 ± 0.07

Table S4.5. Calculated modelfree parameters for 37 nt PBD.

τ_m equals 9.8 ns, and anisotropy of diffusion tensor, D_{ratio} is 1.9

Residue	Spin	S^2	S^2_s	S^2_f	τ_e(ps)	R_{ex}
3	C1’	0.82 ± 0.01	0.82 ± 0.01		1009 ± 304	
4	C1’	0.79 ± 0.01	0.79 ± 0.01		625 ± 186	
7	C1’	0.71 ± 0.02	0.71 ± 0.02		704 ± 190	4. 7 ± 1.4
9	C1’	0.56 ± 0.02	0.56 ± 0.02		709 ± 148	4.6 ± 1.4
11	C1’	0.51 ± 0.02	0.51 ± 0.02		611 ± 114	6.6 ± 1.3
16	C1’	0.67 ± 0.01	0.67 ± 0.01		654 ± 160	
18	C1’	0.80 ± 0.01	0.80 ± 0.01		989 ± 317	
19	C1’	0.75 ± 0.03	0.79 ± 0.02	0.95 ± 0.03	574 ± 155	
20	C1’	0.65 ± 0.03	0.73 ± 0.02	0.89 ± 0.03	1069 ± 304	
21	C1’	0.70 ± 0.03	0.76 ± 0.02	0.92 ± 0.03	741 ± 213	
22	C1’	0.75 ± 0.03	0.79 ± 0.02	0.95 ± 0.03	660 ± 183	
23	C1’	0.62 ± 0.03	0.74 ± 0.02	0.85 ± 0.02	852 ± 241	
25	C1’	0.69 ± 0.03	0.74 ± 0.02	0.93 ± 0.03	1073 ± 292	
28	C1’	0.79 ± 0.01	0.79 ± 0.01		1126 ± 305	8.5 ± 1.6
29	C1’	0.73 ± 0.03	0.77 ± 0.02	0.95 ± 0.03	922 ± 289	
31	C1’	0.72 ± 0.01	0.72 ± 0.01		811 ± 223	2.7 ± 1.4
3	C2	0.67 ± 0.03	0.74 ± 0.02	0.90 ± 0.03	857 ± 278	
12	C2	0.65 ± 0.03	0.72 ± 0.02	0.90 ± 0.03	755 ± 254	
4	C5	0.73 ± 0.01	0.73 ± 0.01		1253 ± 288	
7	C5	0.61 ± 0.03	0.68 ± 0.02	0.90 ± 0.03	724 ± 211	
8	C5	0.64 ± 0.03	0.69 ± 0.02	0.93 ± 0.02	1287 ± 260	
10	C5	0.32 ± 0.02	0.39 ± 0.02	0.81 ± 0.03	482 ± 70	
11	C5	0.41 ± 0.02	0.50 ± 0.02	0.82 ± 0.03	569 ± 101	
13	C5	0.47 ± 0.02	0.58 ± 0.02	0.81 ± 0.02	734 ± 189	
15	C5	0.69 ± 0.03	0.74 ± 0.02	0.93 ± 0.03	1204 ± 302	
16	C5	0.70 ± 0.03	0.75 ± 0.02	0.94 ± 0.03	1268 ± 275	
17	C5	0.71 ± 0.03	0.74 ± 0.02	0.96 ± 0.03	972 ± 317	
20	C5	0.44 ± 0.02	0.55 ± 0.02	0.80 ± 0.02	909 ± 263	
21	C5	0.51 ± 0.02	0.68 ± 0.02	0.75 ± 0.02	742 ± 218	
22	C5	0.71 ± 0.03	0.76 ± 0.02	0.93 ± 0.03	984 ± 314	

25	C5	0.73 ± 0.03	0.77 ± 0.02	0.95 ± 0.03	1107 ± 325
29	C5	0.78 ± 0.01	0.78 ± 0.01		752 ± 268
30	C5	0.59 ± 0.03	0.67 ± 0.02	0.88 ± 0.03	687 ± 188
32	C5	0.74 ± 0.03	0.77 ± 0.02	0.96 ± 0.02	1076 ± 325
33	C5	0.79 ± 0.01	0.79 ± 0.01		1263 ± 290
35	C5	0.68 ± 0.03	0.72 ± 0.02	0.95 ± 0.03	929 ± 297
36	C5	0.52 ± 0.02	0.58 ± 0.02	0.89 ± 0.03	605 ± 119
37	C5	0.68 ± 0.03	0.71 ± 0.02	0.96 ± 0.03	503 ± 113
2	C8	0.64 ± 0.03	0.74 ± 0.02	0.86 ± 0.03	978 ± 305
3	C8	0.66 ± 0.03	0.73 ± 0.02	0.90 ± 0.03	1144 ± 313
5	C8	0.64 ± 0.03	0.73 ± 0.02	0.88 ± 0.02	1054 ± 317
12	C8	0.55 ± 0.02	0.67 ± 0.02	0.82 ± 0.03	682 ± 186
14	C8	0.62 ± 0.03	0.72 ± 0.02	0.85 ± 0.02	1001 ± 291
18	C8	0.69 ± 0.03	0.77 ± 0.02	0.89 ± 0.03	1013 ± 326
23	C8	0.59 ± 0.03	0.73 ± 0.02	0.81 ± 0.03	843 ± 266
24	C8	0.67 ± 0.03	0.76 ± 0.02	0.88 ± 0.03	776 ± 257
27	C8	0.74 ± 0.03	0.80 ± 0.02	0.93 ± 0.03	897 ± 316
28	C8	0.74 ± 0.03	0.82 ± 0.02	0.91 ± 0.03	955 ± 338
31	C8	0.69 ± 0.03	0.77 ± 0.02	0.89 ± 0.03	913 ± 306
34	C8	0.78 ± 0.01	0.78 ± 0.01		593 ± 188

Table S4.6. Calculated S^2 of carbons from ensemble of 62 nt. duck epsilon, FED

Residue	S^2_{calc}(C1’)	S^2_{calc}(C2)	S^2_{calc}(C5)	S^2_{calc}(C6/8)	Residue	S^2_{calc}(C1’)	S^2_{calc}(C2)	S^2_{calc}(C5)	S^2_{calc}(C6/8)
1	0.91			0.95	32	0.97		0.96	0.96
2	0.91			0.95	33	0.90			0.89
3	0.58		0.42	0.82	34	0.88		0.82	0.76
4	0.93			0.95	35	0.93		0.95	0.95
5	0.97		0.96	0.97	36	0.96			0.97
6	0.98		0.96	0.96	37	0.96		0.97	0.97
7	0.97	0.98		0.96	38	0.97		0.97	0.96
8	0.96	0.97		0.95	39	0.97			0.97
9	0.95	0.94		0.93	40	0.97		0.97	0.98
10	0.94			0.94	41	0.98			0.97
11	0.94			0.96	42	0.99		0.98	0.98
12	0.89		0.94	0.92	43	0.99			0.98
13	0.93	0.95		0.95	44	0.58		0.51	0.46
14	0.31		0.32	0.38	45	0.95			0.96
15	0.21		0.42	0.14	46	0.98	0.94		0.97
16	0.49		0.37	0.97	47	0.96		0.96	0.96
17	0.16		0.34	0.11	48	0.49		0.46	0.64
18	0.09		0.08	0.27	49	0.47			0.36
19	0.33	0.29		0.21	50	0.96		0.96	0.96
20	0.18		0.34	0.23	51	0.88	0.94		0.93
21	0.90			0.97	52	0.97		0.95	0.95
22	0.96		0.96	0.97	53	0.96		0.95	0.94
23	0.97		0.97	0.97	54	0.95		0.95	0.94
24	0.97		0.97	0.97	55	0.94		0.95	0.93
25	0.97	0.97		0.96	56	0.94		0.95	0.95
26	0.98		0.97	0.98	57	0.93			0.95
27	0.99	0.97		0.98	58	0.94			0.96
28	0.98		0.98	0.98	59	0.93		0.95	0.95
29	0.94		0.97	0.92	60	0.79	0.89		0.82
30	0.98			0.97	61	0.96		0.96	0.95
31	0.98		0.97	0.97	62	0.97		0.96	0.95

Table S4.7. Structure statistics

Type of restraint	62 nt epsilon duck RMSD	37 nt PBD RMSD
NOE	0.07 ± 0.00 Å	0.04 ± 0.00 Å
RDC	3.37 ± 0.14 Hz	3.44 ± 0.09 Hz
Chem. shift	0.23 ± 0.01 ppm	0.23 ± 0.01 ppm
dihedral angles	0.04 ± 0.01 °	0.06 ± 0.00 °
bonds	0.01 ± 0.00 Å	0.01 ± 0.00 Å
bond angles	1.13 ± 0.01 °	1.72 ± 0.02 °
improper dihedrals	1.00 ± 0.03 °	1.61 ± 0.06 °
structure RMSD	4.06 ± 1.09 Å	0.32 ± 0.14 Å
pairwise RMSD	3.44 ± 1.02 Å	1.94 ± 0.54 Å

References and notes

1. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35:2800-11
2. Flodell S, Petersen M, Girard FC, Zdunek J, Kidd-Ljunggren K, et al. 2006. Solution structure of the apical stem–loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res* 34:4449-57
3. Petzold K, Duchart E, Flodell S, Larsson G, Kidd-Ljunggren K, et al. 2007. Conserved nucleotides in an RNA essential for hepatitis B virus replication show distinct mobility patterns. *Nucleic Acids Res* 35:6854-61
4. Werf RMvd, Girard FC, Nelissen F, Tessari M, Wijmenga SS. 2008. ^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon. *Biomol NMR Assignments* 2:143-5
5. Ampt KAM, Ottink OM, Girard FC, Nelissen F, Tessari M, Wijmenga SS. 2008. ^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV apical stem loop of the epsilon encapsidation signal. *Biomol NMR Assignments* 2:159-62
6. Ampt KAM, Werf RMvd, Wijmenga SS, Tessari M. *Manuscript in preparation*
7. Gelbin A, Schneider B, Clowney L, Hsieh SH, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Am Chem Soc* 118:519-29
8. Allain FHT, Varani G. 1995. Structure of the P1 Helix from Group I Self-splicing introns. *J Mol Biol* 250:333-53
9. Flodell S, Schleucher J, Cromsigt JAMTC, Ippel H, Kidd-Ljunggren K, Wijmenga SS. 2002. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res* 30:4803-11
10. Bailor MH, Musselman C, Hansen AL, Gulati K, Patel DJ, Al-Hashimi HM. 2007. Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat. Protoc.* 2:1536-46
11. Lilley DMJ, Clegg RM, Diekmann S, Seeman NC, Kitzing Ev, Hagerman P. 1995. A nomenclature of junctions and branchpoints in nucleic acids. *Eur.J.Biochem.* 230:1-2
12. Lilley DMJ, Clegg RM, Diekmann S, Seeman NC, Kitzing Ev, Hagerman P. 1996. A Nomenclature of Junctions and Branchpoints in Nucleic Acids. Recommendations 1994. *J Mol Biol* 255:554-5
13. Lilley DMJ. 2000. Structures of helical junctions in nucleic acids. *Quart. Rev. Biophys.* 33:109-59
14. Lilley DMJ. 1998. Folding of Branched RNA Species. *Biopolymers* 48:101-12
15. Bailor MH, Sun X, Al-Hashimi HM. 2010. Topology Links RNA Secondary Structure with Global Conformation, Dynamics, and Adaptation. *Science* 327:202-6
16. Stelzer AC, Frank AT, Bailor MH, Andricioaei I, Al-Hashimi HM. 2009. Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. *Methods* 49:167-73

17. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, et al. 1989. Definitions and Nomenclature of Nucleic Acid Structure Parameters. *J Mol Biol* 205:787-91

18. Yang H, Jossinet F, Leontis NB, Chen L, Westbrook J, et al. 2003. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* 31:3450–60

19. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA, et al. 2008. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320:1471-5

20. El Hassan MA, Calladine CR. 1995. The Assessment of the Geometry of Dinucleotide Steps in Double-Helical DNA; a New Local Calculation Scheme. *J Mol Biol* 251:648-64

21. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37

22. Rosenberg JM, Seeman. NC, Day RO, Rich A. 1976. RNA double helices generated from crystal structures of double helical dinucleotide phosphates. *Biochem Biophys Res Commun* 69:979-87

23. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-21

24. Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* 10:337-50

25. Varani G, Cheong C, Tinoco IJ. 1991. Structure of an Unusually Stable RNA Hairpin. *Biochemistry* 30:3280-9

26. Brunger AT. 1992. *X-PLOR version 3.1 A system for X-ray Crystallography and NMR*. New Haven: Yale University Press

27. Wu B, Girard FC, Buuren BNMv, Schleucher J, Tessari M, Wijmenga SS. 2004. Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nucleic Acids Res* 32:3228-39

28. Wu B, Petersen M, Girard FC, Tessari M, Wijmenga SS. 2006. Prediction of molecular alignment of nucleic acids in aligned media. *J Biomol NMR* 35:103-15

29. Buuren BNMv, Schleucher J, Wittmann V, Griesinger C, Schwalbe H, Wijmenga SS. 2004. NMR Spectroscopic Determination of the Solution Structure of a Branched Nucleic Acid from Residual Dipolar Couplings by Using Isotopically Labeled Nucleotides. *Angew Chem Int Ed* 43:187-92

30. Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman M. 1996. Geometric Parameters in Nucleic Acids: Nitrogenous Bases. *J Am Chem Soc* 118:509-18

31. Feng Z, Westbrook J, Berman HM. 1998. NUCheck. *Rep. NDB-407*, Rutgers University, New Brunswick

32. Ampt KAM, Werf RMvd, Nelissen FHT, Tessari M, Wijmenga SS. 2009. The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding. *Biochemistry* 48:10499-508

33. Dosset P, Hus JC, Marion D, Blackledge M. 2001. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J.Biomol.NMR* 20:223-31

34. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450:1263-7

35. Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY. 2001. NMR studies of Brownian tumbling and internal motions in proteins. *Prog NMR Spectrosc* 38 197-266

36. Korzhnev DM, Skrynnikov NR, Millet O, Torchia DA, Kay LE. 2002. An NMR Experiment for the Accurate Measurement of Heteronuclear Spin-Lock Relaxation Rates. *J Am Chem Soc* 124:10743-53

37. Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, et al. 1994. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* 4:301-6

38. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-93

39. Loria JP, Rance M, Palmer AG. 1999. A Relaxation-Compensated Carr−Purcell−Meiboom−Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J Am Chem Soc* 121:2331-2

40. Mulder FAA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. 2001. Studying excited states of proteins by NMR spectroscopy. *Nature Struct.Biol.* 8:932-5

41. Stueber D, Grant DM. 2002. ^{13}C and ^{15}N Chemical Shift Tensors in Adenosine, Guanosine Dihydrate, 2'-Deoxythymidine, and Cytidine. *J Am Chem Soc* 124:10539-51

42. Duchart E, Schwalbe H. 2005. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR ^{13}C relaxation. *J Biomol NMR* 32:295-308

43. Torre JGdl, Huertas ML, Carrasco B. 2000. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J Magn Reson* 147:138-46

44. Mandel AM, Akke M, Palmer AG. 1995. Backbone Dynamics of Escherichia coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J Mol Biol* 246:144-63

45. Palmer AG, Rance M, Wright PE. 1991. Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance 3C Heteronuclear NMR Spectroscopy. *J Am Chem Soc* 113:4371-80

46. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. 1990. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J Am Chem Soc* 112:4989-91

47. Best RB, Clarke J, Karplus M. 2005. What Contributions to Protein Side-chain Dynamics are Probed by NMR Experiments? A Molecular Dynamics Simulation Analysis. *J Mol Biol* 349:185-203

48. Lipari G. 1982. Protein Dynamics and NMR Relaxation: Comparison of Simulations with Experiment. *Nature* 300:197-8

49. Henry ER, Szabo A. 1985. Influence of vibrational motion on solid state lineshapes and NMR relaxation. *J.Chem.Phys.* 82:4753-61

Chapter 5

Nucleic Acid Structure Determination from NMR Proton Chemical Shifts

Werf, R.M. van der; Tessari, M.; Wijmenga, S.S.: *Manuscript under Revision*

5 Nucleic Acid Structure Determination from NMR Proton Chemical Shifts

5.1 Abstract

We present a method for *de novo* derivation of the three-dimensional helix structure of nucleic acids using non-exchangeable proton chemical shifts as sole source of experimental restraints. The method is called CHEOPS (**che**mical shift *de nov***o** structure derivation **p**rotocol employing **s**ingular value decomposition) and uses iterative singular value decomposition to optimize the structure in helix parameter space. The correct performance of CHEOPS and its range of application are established via an extensive set of structure derivations using either simulated or experimental chemical shifts as input. The simulated input data are used to assess in a defined manner the effect of errors or limitations in the input data on the derived structures. We find that the RNA helix parameters can be determined with high accuracy. We finally demonstrate via two deposited RNA structures that experimental proton chemical shifts suffice to derive RNA helix structures with high precision and accuracy. CHEOPS provides, subject to further development, new directions for high-resolution NMR structure determination of nucleic acids.

5.2 Introduction

Structure determination of nucleic acids from NMR data has been an intrinsically difficult problem for many years(1-6). This is mainly due to the extended nature of helices, which conflicts with the short-range structure information contained in common NMR parameters, like NOEs and J-couplings, while severe spectral overlap, low proton densities, spin-diffusion(7; 8), and dynamics(9; 10) add to these difficulties. The introduction of heteronuclear residual dipolar couplings (RDCs) has helped to alleviate these problems, in part, thanks to the global structure information contained in these parameters(11; 12). Unfortunately, commonly used heteronuclear RDCs only carry orientation but no translational information(13). As a result of this incompleteness of NMR experimental restraints, NMR structures may still contain inaccuracies, e.g. in helical rise(14; 15) and non-bonded terms in the force field can significantly affect the derived structures(1; 16). Introduction of other more complete and structure-sensitive restraints is thus of relevance.

Chemical shifts are exquisitely sensitive probes of (bio)molecular structure. They provide detailed information on structure and electronic properties in solution, non-crystalline, and crystalline states(17-24). For proteins, chemical shifts are commonly employed to establish secondary structure and backbone torsion angles(17; 25). Recently, it has even been demonstrated that the three-dimensional structure of a protein can be derived, using chemical shifts as sole source of experimental restraints(18; 21-24). These methods make use of molecular replacement, which in turn is based on sequence similarity and empirical structure-chemical shift relationships established from analyses of large databases of protein structures and their related chemical shifts. In contrast, for nucleic acids chemical shifts have until now been less extensively employed to derive structure information. This may in part be due to the fact that precise and accurate structure-chemical-shift relationships for heteronuclear shifts are limited for nucleic acids(17; 26). On the other hand and in contrast to proteins, well-parameterized ring-current and magnetic-anisotropy effects govern proton chemical shifts of nucleic acids. The associated semi-classical equations provide a way to precisely and accurately predict non-exchangeable chemical shifts given a three-dimensional structure(19; 20; 27), e.g. the RMSD between predicted and observed proton chemical shifts in well-defined rigid RNA helices is found to be as small as 0.08 ppm(19). This approaches the precision obtained from purely empirical relationships(17; 26; 28). In contrast to heteronuclear RDCs, the semi-classical equations for ring current and magnetic anisotropy effects contain information on both translational and orientational helix parameters. Together with their high sensitivity to structure parameters, evident from the semi-classical equations, these characteristics suggest that proton chemical shift restraints may alleviate at least part of the problems associated with NMR RNA/DNA structure calculations.

Here we describe how proton chemical shifts can be used to *de-novo* derive the NMR structures of nucleic acid helices at high resolution without the requirement of any additional experimental parameters. The method is called CHEOPS (**che**mical shift *de nov***o** structure derivation **p**rotocol employing **s**ingular value decomposition and backbone restrained molecular dynamics). Simulated input data are used to assess in a defined manner the effect of errors or limitations in input data on the derived structures, while the structure calculations on two published RNA structures employing experimental input chemical shifts demonstrate that CHEOPS also works in practice.

5.3 Materials and methods

Allowed helix space and ‘standard’ helix in terms of helix parameters

The helix parameters were analyzed, using 3DNA v1.5(29), in 20 RNA helix structures solved by X-ray diffraction to a resolution better than 3 Å (157D, 1DQH, 1I9X, 1KFO, 1RNA, 1RXA, 1RXB, 1SDR, 1ZX7, 205D, 255D, 259D, 353D, 377D, 402D, 405D, 413D, 433D, 438D, 472D). The average helix parameters and their standard deviations (σhelixpar) are listed in Table S5.1. The allowed A-helix space is defined by average helix parameter values +/- 3σhelixpar. Likewise, the allowed B-helix space (not used here) can be derived from the values given in Table S5.1.

Evaluating CHEOPS using simulated chemical shifts

A fiber-model A-helix, as defined by Olson et al.(14) and built using 3DNA(29), was taken as target and called the *target structure*. It consisted of six canonical base pairs with primary sequence 5’-ACAGCU-3’:5’-AGCUGU-3’. A set of 93 ^{1}H chemical shifts, i.e. H1’, H2’, H3’, H4’, H5’, H5”, H2, H5, H6, and H8, was calculated with NUCHEMICS(20) using the Giessner-Prettre parameters set (GP set)(30; 31) from this *target structure*. These simulated chemical shifts were considered as ‘observed shifts’ in the test structure derivations using CHEOPS (see Results, CHEOPS structure calculation protocol). In these test calculations, chemical shifts were calculated by means of NUCHEMICS using the GP parameter set, except otherwise stated. As starting structures, 1000 helices were generated with each helix parameter randomly selected in the interval $[-3\sigma_{\text{helixpar}}, 3\sigma_{\text{helixpar}}]$. Further, the threshold (parameter λ) for zeroing small eigenvalues of the G-matrix was set to a value of 0.1%; as discussed in supplementary material, λ cannot be set to zero because then the system-imprecision would introduce large fluctuations into the solution $\mathbf{\Delta H}$ vector. The parameter κ, that scales the solution vector, was set to 0.50. The maximum number of cycles in the iterative SVD was set to 150. Four main sets of test calculations were carried out with the convergence tolerance set to 0.05, 0.10, 0.15, and 0.30 ppm, respectively.

The converged SVD structures were checked for holonomic(32) and standard hydrogen bond distance(33) violations in Xplor v3.851(34). Violating structures were not considered further. Because the converged SVD structures calculated with a complete set of chemical shifts rarely show such violations, no ‘backbone regularization’ (see supplementary material) was applied to the converged SVD structures. The structures that pass all tests are called accepted structures.

Each accepted structure (and/or converged SVD structure) was compared to the *target structure* in terms of helix parameters as well as atomic coordinates. The average pairwise RMSD ($\mathrm{RMSD}_{\mathrm{pairwise}}$) of the atomic coordinates of the ensemble of accepted structures (and or converged SVD structures) was calculated; it is a measure of the precision of the protocol. In addition, the average accepted structure (and/or converged SVD structure) was built using 3DNA based on the average helix parameters of the entire ensemble. The RMSD of this average structure to the *target structure*, $\mathrm{RMSD}_{\mathrm{av2target}}$, is a measure of the accuracy of the protocol.

To estimate the effect of uncertainty in parameterization of the ring current and magnetic anisotropy on the chemical shift derived structures the foregoing test calculations were repeated, using the more recent parameter set derived by Case(35) (DC set; see supplementary material for details).

Effect of the average number of chemical shift restraints per base pair. The effect of a smaller number of chemical shifts restraints on the precision and accuracy of the derived structures was tested on a fiber-model A-helix of 10 canonical base pairs and primary sequence 5’-ACGUGCGUAC-3’:5’-UGCACGCAU-3’. For this helix, a set of 154 chemical shifts of non-exchangeable protons was calculated with NUCHEMICS(19; 20), using the GP-parameter set. From this set, 300 new sets of ‘observed’ shifts were generated as follows. First, 200 sets were obtained by randomly removing 4, 14, 24, 34, 44, 54, 64, 74 chemical shifts with 25 subsets per deletion. For each of these 200 sets, the structure derivation was run starting with 20 random structures ($\pm 3\sigma_{\mathrm{helixpar}}$). This corresponds to 500 structure calculations per random deletion subset (4000 in total). The remaining 100 sets of chemical shifts were generated by randomly removing 84, 94, 104, 114, 124 chemical shifts with 20 subsets per deletion. For each of these 100 sets, the structure derivation was run starting with 250 random A-helices. This corresponds to 5000 structure calculations per random deletion subset. In addition, a set of helices was calculated employing all ‘observed’ chemical shifts, starting from 500 random A-helices. Finally, three separate sets of experimentally ‘easily’ accessible chemical shifts were generated: a complete list of aromatic proton shifts complemented with H1’/2’ shifts (set 1) or H1’/4’ shifts (set 2) or H1’ shifts (set 3). These calculations started from 500 random A-helices. The parameters λ and κ were set as before and the chemical shift tolerance to 0.10 ppm. The converged SVD structures were checked for hydrogen bond violations in the basepairs and subsequently backbone regularized (see supplementary material).

The accepted structures were analyzed as before. In addition, the quality of the accepted structures was assessed by means of a Validation Chemical Shift Q-factor (VCSQ-factor) defined as:

$$VCSQ = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\delta_{conf,calc,i} - \delta_{conf,obs,i}\right)^2}{\frac{1}{N}\sum_{j=1}^{N}\delta^2_{conf,obs,j}}} \tag{5.1}$$

Here, N equals the total number of non-exchangeable proton chemical shifts. The VCSQ-factor is a measure of how well the resulting structure explains the complete set of proton chemical shifts.

Evaluating CHEOPS on nucleic acids of known structure using experimental chemical shifts

Two NMR RNA hairpin structures with canonical basepairs in the A-helix stem were retrieved from the PDB(http://www.pdb.org) and their matching proton chemical shifts from the BMRB (http: //www.bmrb.wisc.edu): 1PJY (BMRB-entry:5834) and 2JTP (BMRB-entry:15417). The CHEOPS protocol used for simulated input data was followed in broad lines, except for the following changes. 1) The input chemical shifts are now the real experimental chemical shifts. They were checked for outliers or incorrect calibrations by running the iterative SVD protocol using a small number (~ 5) of randomized starting helices and parameter settings as described below for the full protocol. Experimental chemical shifts that consistently deviated more than the chemical shift tolerance from the calculated values were provisionally marked as outliers. They were only removed when an acceptable reason for removal did exist (see results for details). 2) The structure calculations started from a set of 500 random starting helix structures, generated within A-helix space with a variation of $\pm 3\sigma_{\mathrm{helixpar}}$ around the standard A-helix values. 3) The parameter settings for the iterative SVD were adjusted compared to the CHEOPS protocol for simulation data. The λ-value was set to 4.25%, a value optimal for providing stable SVD solutions given the error in the experimental chemical shift of a few percent (measurement error plus estimated back-calculation error). The scaling factor was set to 0.5 and the maximum number of cycles in the iterative SVD to 150.

5.4 Results and discussion

CHEOPS structure calculation protocol

The formal basis for CHEOPS, the proposed method for nucleic acid helix structure calculation from chemical shifts, is presented below. We employ the helix parameter description(15), because it provides a quantitative framework for describing the conformation of nucleic acids, and thus for understanding the, sometimes subtle, variations in their three-dimensional structures(14; 36). Moreover, three-dimensional atomic models can be reconstructed given a set of helix parameters(29) and the reduced space spanned by the helix parameters (12 per basepair) allows one to focus on the essential conformational features rather than having to consider all degrees of freedom, as would be the case when nucleic acid conformation is described in terms of atomic coordinates. In the CHEOPS structure derivation, we use the program 3DNA and its up-to-date helix parameter analysis and atomic-model rebuilding facilities(29), while the proton chemical shifts are calculated given a three-dimensional structure using NUCHEMICS(19; 20) and are thus based on ring current and magnetic anisotropy as discussed below.

The conformational chemical shift of a proton k in a helix ($\delta_{conf,k}$) can be expressed as a function g of the complete set of helix parameters that describe the (helix) structure,

$$\delta_{conf,k} = g\left(S_{x,1}, S_{y,1}, S_{z,1}, \kappa_1, \omega_1, \sigma_1, dx_1, dy_1, D_{z,1}, \eta_1, \theta_1, \Omega_1, \cdots, S_{x,n}, S_{y,n}, S_{z,n}, \kappa_n, \omega_n, \sigma_n, dx_n, dy_n, D_{z,n}, \eta_n, \theta_n, \Omega_n\right) \tag{5.2}$$

The change in proton chemical shift, $\Delta\delta_{conf,k}$, from its current value $\delta^{o}_{conf,k}$ resulting from a change of one or more helix parameters can be calculated by a Taylor expansion of the function g. For small changes such a series can be truncated to first order leading to equation (5.3).

$$\Delta\delta_{conf,k} \approx \sum_{basepair}\sum_{helixpar}\left(\frac{\partial\delta_{conf,k}}{\partial H_{basepair,helixpar}}\right)_{H=H^{o}_{basepair,helixpar}} \Delta H_{basepair,helixpar} \tag{5.3}$$

Here, $\Delta H_{basepair,helixpar}$ is the change in helix parameter $H_{basepair,helixpar}$ from its current value $H_{basepair,helixpar}{}^{o}$. The partial derivatives in equation(5.3) can be evaluated numerically by reconstructing via 3DNA an atomic model from the helix parameters and employing NUCHEMICS(19; 20) and thus the semi-classical equations for the ring current and magnetic anisotropy effect that govern the proton conformational chemical shifts of nucleic acids. In compact notation, equation(5.4) provides an expression for the change in proton chemical shifts upon a (small) change in the initial helix structure:

$$\Delta\mathbf{C} = \mathbf{G}\cdot\Delta\mathbf{H} \tag{5.4}$$

Here, $\mathbf{\Delta C}$ (=$\mathbf{C}$-$\mathbf{C_o}$) and $\mathbf{\Delta H}$ (=$\mathbf{H}$-$\mathbf{H_o}$) are column vectors containing the changes in conformational proton chemical shifts and helix parameters, respectively, while **G** is the matrix whose elements contain the partial derivatives of the chemical shift versus helix parameters and is evaluated numerically (see above and supplementary material). In equation(5.4), the helix parameters in **ΔH** and **G** are expressed in units of σ_{helixpar} (Table S5.1). The position of the flanking residues (i-1) and (i+1) and to a smaller degree (i-2) and (i+2) affect the proton chemical shifts of residue i. Consequently, the G-matrix has a block-diagonal form. The aim is to solve equation(5.4) given a list of chemical shift differences (**ΔC**):

$$\Delta\mathbf{H} = \mathbf{G}^{-1}\bullet\Delta\mathbf{C} \tag{5.5}$$

When **ΔC** is the difference between calculated and observed shifts, **ΔH** contains changes required in helix parameters to obtain the structure consistent with observed chemical shifts. The required inverse of the G-matrix in equation (5.5) can be obtained via singular value decomposition (SVD). The new helix parameters, $\mathbf{H_{new}}$, are subsequently calculated from **ΔH**,

$$\mathbf{H}_{new} = \mathbf{H_0} + \kappa\cdot\Delta\mathbf{H} \tag{5.6}$$

Here, $\mathbf{H_0}$ is the vector that contains the previous helix parameters. The factor κ is a user chosen parameter (default value=0.50). Because proton chemical shifts and helix parameters are often close to but not completely linearly related (Figure 5.1a-d), an iterative procedure is therefore followed to find the final solution. A small value for the factor κ assures that, in the iterative procedure, subsequent solution vectors **ΔH** follow the potential curvature of the helix

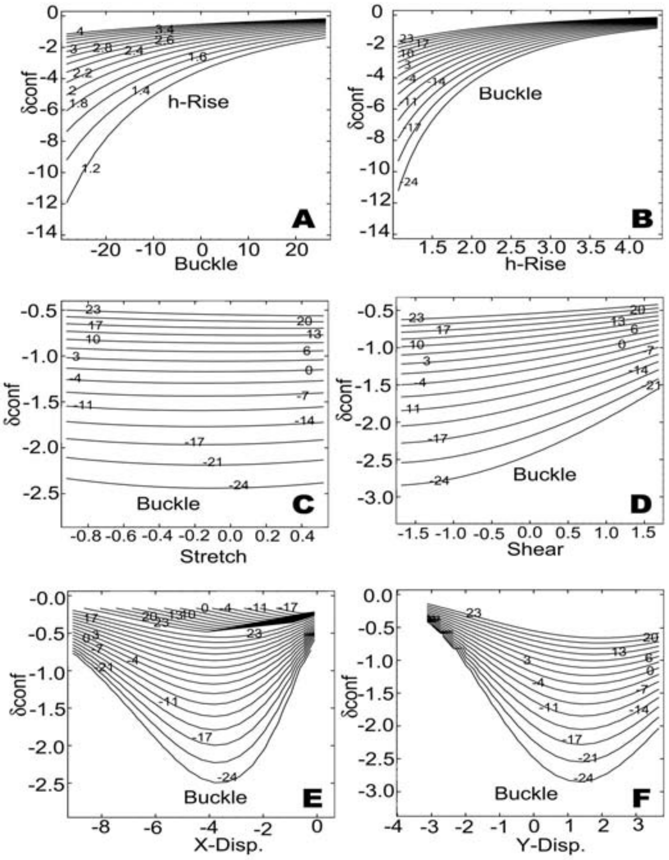


Figure 5.1: Conformational chemical shift of the H5 proton, δ_{confH5}, of residue C4 in a standard anti-parallel A-helix with 6 base pairs and primary sequence 5'-GGACGA-3':5'-UCGUCC-3', as function of the helix parameters of basepair 4 (C4:G9). The helix parameters of basepair 4 were varied in the interval [$-4\sigma_{helixpar}$, $4\sigma_{helixpar}$], while keeping the other values fixed at the standard A-helix values (Table S5.1). For each combination of helix parameters an atomic model was built with the rebuilding feature in 3DNA(29) and its proton chemical shifts calculated with NUCHEMICS(20). The number in the contour lines indicates the value of the second helix parameter that is varied. (a, b) Typical example of the usual monotonic dependence of chemical shift upon change in helix parameter is shown; a) δ_{confH5} vs. Buckle (h-rise); b) δ_{confH5} vs. h-Rise (Buckle). (c, d) Example is shown of helix parameters, that barely affect the chemical shift; c) δ_{confH5} vs. Stretch (Buckle); d) δ_{confH5} vs. Shear (Buckle). (e, f) Example is shown of helix parameters with a parabolic dependence on chemical shift; e) δ_{confH5} vs. X-disp (Buckle); f) δ_{confH5} vs. Y-disp (Buckle). A complete analysis using the partial derivatives of the proton chemical shift with respect to the helix parameters shows that the sensitivity of chemical shifts to structural changes follows the trend: H5≈H2>H6≈H2'≈H8>H1'≈H3'> H5'≈H4'>H5". Alternatively, the sensitivity of helix parameters to changes in the chemical shift is given by: *buckle > Y-displacement ≈ tip>propeller twist>inclination ≈ rise> helix-twist ≈ X-displacement>stagger>opening> shear>stretch*

parameter-chemical shift space. In each step (cycle) in the iterative procedure an atomic model of the new helix is rebuild from $\mathbf{H}_{\mathbf{new}}$ using 3DNA(29), from which new chemical shifts are calculated using NUCHEMICS(19; 20). A new cycle can then start. The scheme continues until calculated and observed shifts are equal within a chemical shift tolerance set by the user. The obtained structures are called converged SVD structures. As a final step, the converged SVD structures undergo backbone regularization, when needed. For this, a short restrained molecular dynamics (rMD) run is executed in Xplor(34), during which the aromatic planes are kept fixed in Cartesian space and loose standard A-helix distance and/or sugar-backbone torsion angle restraints are employed. In this way, the global structure and thus helix parameters remain unchanged during backbone regularization (see supplementary material for further details). The above forms the basis for the calculation of a nucleic acid helix structure from chemical shifts presented here. Figure 5.2 shows the complete protocol of CHEOPS.

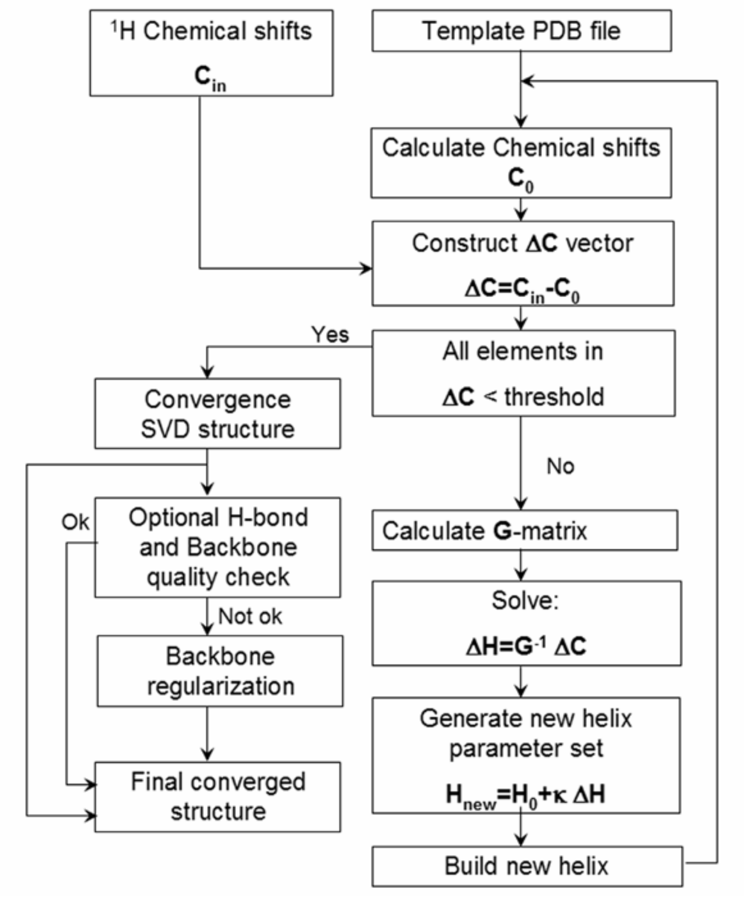


Figure 5.2: Flowchart of the protocol for the derivation of helical nucleic acid structures from chemical shifts. The template PDB file is read, and chemical shifts are calculated from this structure. The calculated chemical shifts are subtracted from the restraint chemical shifts, thereby composing vector ΔC. Subsequently, convergence is judged, i.e. are all chemical shifts within a user-set threshold limit? If ‘Yes’, the converged SVD structure is checked from compliance backbone and h-bond parameters, its backbone may be regularized when required by means of a brief restrained molecular dynamics runs during which the bases are kept fixed and subsequently the final converged structure is obtained. If ‘No’, the ΔH vector is solved from the G-matrix and the chemical shift deviation (ΔC) is solved and a new set of helix parameters calculated H_{new}. The ΔH vector is scaled by κ in this calculation, to prevent too large changes and thereby restraining the solution to the regime, where the approximation of a linear relationship between chemical shifts and helix parameters holds. From the new helix parameter set, H_{new}, a new atomic model is built, and the calculation enters its new cycle.

A few aspects should be stressed. The CHEOPS protocol starts from a set of random starting structures that well cover allowed A-helix conformational space: a space defined by helices that are within +/-3σ from average helix parameter values, as follows from an analysis of a representative set of helices. In this way, no bias is introduced in the set of final accepted structures. There are two situations when bias may otherwise arise. 1) If the number chemical shifts is smaller than the number of helix parameters the SVD solution may depend on the starting structure (see supplementary material for a more extensive description of this point). On the other hand, if the number of chemical shifts is larger than the number of helix parameters, the SVD solution forms a least-squares solution and does not depend on the starting structure. 2) Parabolic dependences between chemical shifts and helix parameter (Figure 5.1e,f) might potentially introduce a bias in the set of converged of SVD solutions, because the iterative SVD might not be able to cross the minimum of such curves. However, the set of random starting structures would sample both sides of the *X/Y-displacement* curve as shown in Figure 5.1e,f, thereby removing any bias. However, we find that our iterative calculation protocol crosses the minimum seen for individual chemical shifts like H5 (see Figure S5.2). In other words, this issue does not appear to pose a problem. The likely reason is that each helix parameter depends on a number of chemical shifts of which most show a monotonic and approximately linear dependence.

Evaluating CHEOPS using simulated chemical shifts

The performance of CHEOPS with regard to precision and accuracy, parameterization of ring current and magnetic anisotropy equations, and number of chemical shifts was evaluated as described below. To assess these aspects in a defined manner for essentially all A-helices, simulated chemical shifts were employed and A-helices that cover largely all combinations of canonical basepairs. In addition, A-helices are well established and thus provide a reliable means to test the validity of the chemical shift derived structures. Here we present the main results, while supplementary material contains those of additional tests.

Assessment of precision and accuracy The CHEOPS protocol was run as described in Materials and Methods, starting from 1000 random structures that cover allowed A-helix space (±3$\sigma_{helixpar}$; Figure 5.3, left) with a chemical shift tolerance ranging from 0.05 ppm to 0.30 ppm (Table 5.1) and employing 15 chemical shifts per basepair. We consider the smallest chemical shift tolerance first. Out of the 1000 starting structures, 956 structures

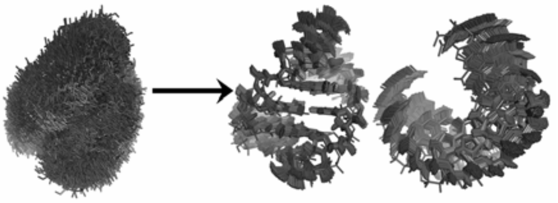


Figure 5.3: . Illustration of the RNA structure-derivation based on chemical shifts via a 6-basepair RNA A-helix with sequence 5’GGACGA3’/5’UCGUCC3’. On the left, overlays are shown of the 1000 random starting structures, which cover ± 3$\sigma_{helixpar}$ of allowed A-helix space. On the right, overlays of the final 956 accepted structures are displayed in two views. A set of restraints consisting of 15 chemical shifts per basepair were used, the threshold for convergence was set to 0.05 ppm, and the GP-parameter set was used in the chemical shift calculations

Table 5.1. Characteristics of the accepted structures resulting from structure derivation of an A-helix based on non-exchanging ^{1}H chemical shifts[&1]

Parameter set[&2]	GP	GP	GP	GP	DC
Chemical shift tolerance[&3]	0.05 ppm	0.10 ppm	0.15 ppm	0.30 ppm	0.10 ppm
N[&4]	956	956	955	963	900
$\mathrm{RMSD}_{\mathrm{av2target}}$ (Å)[&5]	0.33	0.38	0.41	0.44	0.57
$\mathrm{RMSD}_{\mathrm{pairwise}}$ (Å)	0.50	0.68	0.81	1.13	0.84
$\mathrm{RMSD}_{\mathrm{chem.shift}}$ (ppm)	0.011(2)	0.019(5)	0.027(7)	0.055(15)	0.023(3)
<$\mathrm{RMSD}_{\mathrm{helixpar2target}}$>(ppm)[&6]	0.55	0.74	0.88	1.23	0.97

[&1] As ‘observed’ chemical shifts were used the 156 ^{1}H non-exchangeable proton chemical shifts of the 6 basepair A-helix calculated when in its fiber model A-helix conformation (see text) and using the Giessner-Prettre (GP) parameter set for ring-current/magnetic anisotropy. During the iterative SVD structure calculations the ^{1}H chemical shifts were calculated from intermediate structures using either the ring-current/magnetic anisotropy parameter set of Giessner-Prettre (GP) or the ring-current parameter set of Case et al. (DC) (see text). The starting ensemble of 1000 random structures ($\pm 3\sigma_{\mathrm{helixpar}}$) has an $\mathrm{RMSD}_{\mathrm{pairwise}}$ of 5.56 Å and an $\mathrm{RMSD}_{\mathrm{chem.shift}}$ of 1.35 ppm. [&2] GP is the Giessner-Prettre parameter set and DC, Case parameter set. [&3] The chemical shift tolerance for convergence in the iterative SVD. [&4] N is the number of accepted structures. [&5] $\mathrm{RMSD}_{\mathrm{av2target}}$ is the RMSD of the average final accepted structure (see text) to the *target structure* and represents the accuracy. $\mathrm{RMSD}_{\mathrm{pairwise}}$ is the pairwise RMSD of the accepted structures and represents the precision. The $\mathrm{RMSD}_{\mathrm{chem.shift}}$ is the RMSD of the chemical shifts of accepted structures with respect to the ‘observed’ values, while in parentheses is given the standard deviation in $\mathrm{RMSD}_{\mathrm{chem.shift}}$. [&6] The mean RMSD of the helix parameters from *target structure* values normalized to $\sigma_{\mathrm{helixpar}}$; the average is taken over all helix parameters in all bases and structures in the final ensemble**.**

converged; they were all accepted, because none showed extended or compressed hydrogen bonds. In the narrow bundle of overlaid accepted structures the basepairs can clearly be distinguished (Figure 5.3, right). The $\mathrm{RMSD}_{\mathrm{pairwise}}$ of this accepted set is quite small (Table 5.1), indicating that the chemical shifts indeed define the structures with high precision. This is also evident from the RMSDs of the helix parameters of this set, which range from 0.3 to $1.1\sigma_{\mathrm{helixpar}}$ with a mean of $0.55\sigma_{\mathrm{helixpar}}$ (Table 5.1); the translational parameters and rotational parameters are thus defined on average within 0.6 Å and 6°, respectively. In addition, the average converged structure was built from the mean values of the ensemble averaged helix parameters using 3DNA. Note that in this way an average structure is built with correct bond distances and angles. The $\mathrm{RMSD}_{\mathrm{av2target}}$ value (Table 5.1) of this average converged structure versus *target structure* is smaller than the $\mathrm{RMSD}_{\mathrm{pairwise}}$. Thus, the *target structure* lies well within the bundle of converged structures, which shows that no bias is introduced. The RMSD of the helix parameters and $\mathrm{RMSD}_{\mathrm{pairwise}}$ increase linearly with chemical shift tolerance (Table

5.1, Figure S5.5/S5.6). However, at the highest tolerance (0.30 ppm) the precision remains acceptable (mean helix parameter RMSD, $1.23\sigma_{helixpar}$; $RMSD_{pairwise}$, 1.13 Å, Table 5.1). The increase in $RMSD_{av2target}$ is quite small (from 0.33 Å to 0.44 Å), so that $RMSD_{av2target} << RMSD_{pairwise}$. Consequently, the *target structure* remains well within the bundle of accepted structures for all tolerance settings. In other words, higher tolerance does not introduce bias or inaccuracy. We finally note that the variation in the structures resulting from differences in the two best parameterizations of ring current and anisotropy is less than the uncertainty in the derived structures (Table 5.1 and Supplementary Material, Table S5.13/S5.14).

Effect of the number of chemical shift restraints per basepair. We find that down to ~4 chemical shifts per basepair, helix structures can be derived from chemical shifts restraints with reasonable precision ($RMSD_{pairwise} < 1.5$ Å, Table S5.15, Figure S5.7) and accuracy, i.e. no bias because $RMSD_{av2target} << RMSD_{pairwise}$. The VCSQ-factor of the accepted structures stays below 0.5 indicating that even non-employed chemical shifts remain reproduced rather well (Table S5.15, Figure S5.7). Below 4 chemical shifts per basepair the number of converging and physically feasible resulting structures strongly reduces.

We also investigated three data sets of experimentally 'easily' accessible chemical shifts (base protons plus H1'/2' plus H1'/4' or plus H1', Materials and Methods). The VCSQ-factors of these accepted structures (black crosses in Figure S5.7A) reside just below the line of the 'regular' jackknife test results, indicating slightly better correspondence. Similarly, the $RMSD_{pairwise}$ (and $RMSD_{av2target}$) are equal or slightly better (Table S5.15) than the trend of removal of random chemical shifts would suggest (Figure S5.7B). This most likely results from the more even distribution of the chemical shifts over the helix, i.e. they do not cluster in certain residues as may happen in the sets with randomly removed shifts. For all three shift sets the $RMSD_{pairwise}$ is ca. 1 Å, which indicates that per basepair aromatic proton shifts in combination with at least one or two sugar proton shifts (H1' and/or H2'/4') suffice to accurately define the helix structure.

Evaluating CHEOPS on nucleic acids of known structure using experimental chemical shifts

Strategy and selection/analysis of pdb structures. To evaluate the performance of CHEOPS using experimental chemical shifts the helical stem structures of two published RNA stem-loops were derived *de novo* from the available proton chemical shifts, namely the HIV-1

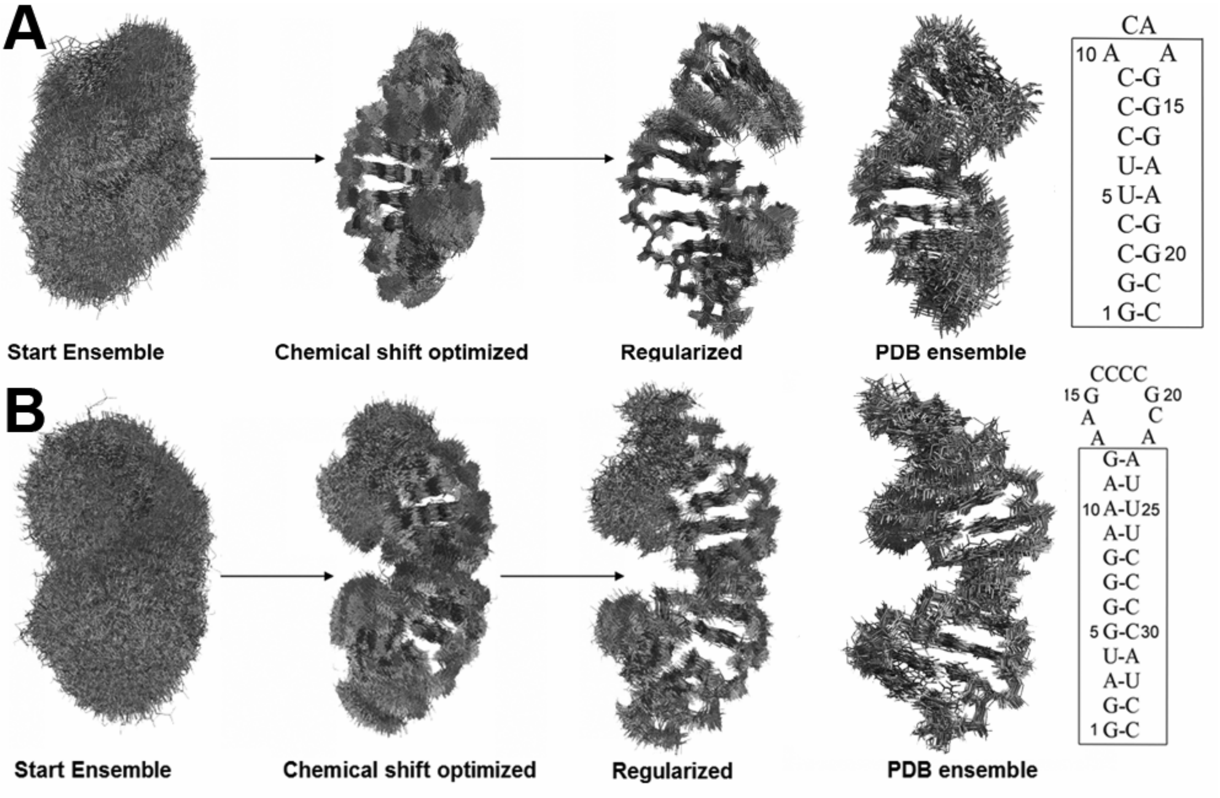


Figure 5.4: Structure derivation using experimental proton chemical shifts demonstrated on the HIV-1 (A) and SIV (B) frame-shift inducing RNA elements. Sequence and secondary structures are shown to the far right. The boxed regions contain the helical segments whose structures were derived using observed chemical shifts as experimental restraints. For the HIV-1 RNA, the conformation of the non-canonical A:A basepair in the tetraloop and of the preceding C:G basepair were kept fixed by keeping their helix parameters at the mean values of the deposited NMR ensemble 1PJY. No basepairs were kept fixed in the SIV1 RNA. The ensemble of starting, converged-SVD, final-accepted structures, PDB-deposited structures are shown from left to right. The chemical shift tolerance was set to 0.15, the parameters λ and κ were set to 4.25 and 0.5, respectively, and the maximum number of iterative SVD cycles was set to 150. For the HIV-1 and SIV RNA elements 87 and 108 chemical shifts were used, respectively. As described in the text, the backbone was regularized by means of a brief restrained molecular dynamics run, using standard helix NOE restraints (for HIV-1 176; for SIV 209) and during which the bases were kept fixed. For SIV, the backbone regularization was for comparison also carried out using 403 published experimental NOE restraints, with or without the published experimental 24 RDCs (Dax = -53.80 Hz, R=0.10).

frame shift inducing RNA stem-loop(37) (HIV_RNA, 1PJY) and the SIV frame shift inducing RNA element(38) (SIV_RNA, 2JTP; see also Materials and Methods). The stem-loop structures are based on standard NMR restraints. No independent X-ray structures are available for these RNAs. However, their helix regions contain only Watson-Crick basepairs (Figure 5.4) and their helix parameters are thus expected to fall within allowed ranges for A-helices. The helix parameters of the stem sections of the pdb structures of HIV1 and SIV RNA indeed all fall within $\pm 3\sigma_{\text{helixpar}}$ of the standard A-helix values (Table S5.16a-b). Deviations between 2-3σ_{helixpar} are found mostly in terminal residues and thus appear to be due to end-effects. Nevertheless, a number of deviations of 2.0-2.3σ_{helixpar} appear in the middle of the helix region. For HIV1 (1PJY), noteworthy are the positive propeller twist values throughout the stem (standard A-helix value -2±7°), as well as two helical rise values that deviate just $2\sigma_{helixpar}$ from the A-helix value of 2.7 Å. For SIV(2JTP), noteworthy are the increased helical rise for three stem residues, a reduced X-displacement and positive propeller twist for two residues. We hypothesize that in the deposited RNA structures, derived with standard NMR methods, such deviations at least in part occur due to limitations in the number and accuracy of NOE-derived distance restraints. We employed 87 and 108 shifts for the stem structures of HIV and SIV, respectively (ca 9.7 shifts/basepair). The tests with simulated data indicate that given ca 9.7 shifts/basepair the $RMSD_{pairwise}$ of accepted chemical shift structures should be ca 1 Å (Figure S5.7) and the RMSD in their helix parameters ca $0.5\text{-}1.5\sigma_{helixpar}$ (Figure S5.5 and S5.6). It thus expected that the larger deviations in the pdb structures should disappear in the chemical shift based structures.

For each of the two RNA stem-loops, the proton chemical shifts of loop residues including the non-canonical closing basepair were not employed as restraints, because end-effects from the residues in the capping loop may affect them. For HIV RNA (1PJY), we removed in addition some H3'/4' shifts, because they could be affected by puckering and/or end-effects (G14 H3', C22 H3'/4') or puckering alone (G15 H3', G19 H3', C21 H3'). These sugar proton shifts were identified as 'outliers' as described in Materials and Methods. For 2JTP (SIV RNA), we removed the H5' sugar proton shifts of G1, G2, A3, U4, G5, and U25 (identified as 'outliers, see Materials and Methods), because they are difficult to assign stereo-specifically and/or affected by end-effects (G1); the removed 'outlier' H1'/2' of C34 could be affected by end-effects. In total, we removed 6/8 out of the 93/116 published proton shifts for the HIV/SIV_RNA stems.

The HIV-1 and SIV frame-shift inducing RNA stem loops (HIV1_RNA and SIV_RNA). For HIV1_RNA (Figures 5.4), the $RMSD_{chem.shift}$ drops from ~4 ppm in the 500 starting structures to 0.06 ppm in the ensemble of converged SVD structures (Table S5.17a), a considerable improvement over the $RMSD_{chem.shift}$ of 0.16 ppm in the deposited NMR structure (Table S5.17b). The width of the structure bundle ($RMSD_{pairwise}$) decreases from 3.3 Å in the starting ensemble to 0.9 Å. For SIV_RNA (Figure 5.4), similar numbers are obtained (Table S5.17b). These converged SVD structures were subsequently backbone regularized and checked for violations (Materials and Methods) to obtain the final accepted structures.

For HIV1-RNA, a relatively small decrease in the $\mathrm{RMSD}_{\mathrm{pairwise}}$ is seen upon backbone regularization in the final accepted structures (from 0.9 Å to 0.7 Å; Tables S5.17a/S5.18a). The $\mathrm{RMSD}_{\mathrm{chem.shift}}$ usually increases somewhat during backbone regularization; here from 0.06 to 0.12 ppm (Table S5.17a/S5.18a). This increase is mostly caused by changes in H1' and H2' proton shifts, resulting from slight reorientations in the corresponding C-H bonds (< 2°) during backbone regularization. Furthermore, the common ribose unit for N- and S-type puckering used in Xplor does not exactly match the individual N- and S-type ribose units used in 3DNA. This aspect contributes additionally to C-H bond reorientation in the sugar and thus to the slight changes in proton chemical shifts. The impact of backbone regularization on the $\mathrm{RMSD}_{\mathrm{chem.shift}}$ illustrates again the high sensitivity of the chemical shift to changes in structure. We finally note that the final accepted structures show no violations (> 1.0 Å) with respect to the standard A-helix inter-residue sugar-sugar and sugar-aromatic distances(2), attesting to the quality of these structures.

For SIV_RNA, the backbone regularization using the standard A-helix NOE restraints, led to an $\mathrm{RMSD}_{\mathrm{chem.shift}}$ of 0.08 ppm in the final accepted structures, essentially the same as in the converged SVD structures. We also performed backbone regularization with various other combinations of experimental and standard A-helix NOE and RDC restraints. The results are comparable in terms of the various RMSDs (Table S5.18b). The published NOE and RDC restraints permitted an independent validation of the final accepted structures. In the final accepted structures, the average $\mathrm{RMSD}_{\mathrm{RDC}}$ ranges from 19 (12 %) to 23 Hz (14 %; Table S5.18b) in the absence of RDC restraints during backbone regularization. It drops to 13 Hz (8 %) in the presence of RDCs (Table S5.19b). The percentages are calculated as the ratio of the RMSD over three times the absolute value of the axial component of –54 Hz of the alignment tensor. These values are just above the $\mathrm{RMSD}_{\mathrm{RDC}}$ of 10 Hz (6%) of the deposited NMR structure set. The reduction in $\mathrm{RMSD}_{\mathrm{RDC}}$ upon inclusion of RDC restraints in the backbone regularization is due to sugar C-H RDC restraints that optimize sugar conformation, because the bases are kept fixed during backbone regularization. The final accepted structures do not show any distance violations > 1.5 Å with respect to the BMRB experimental NOE distance restraints(38).

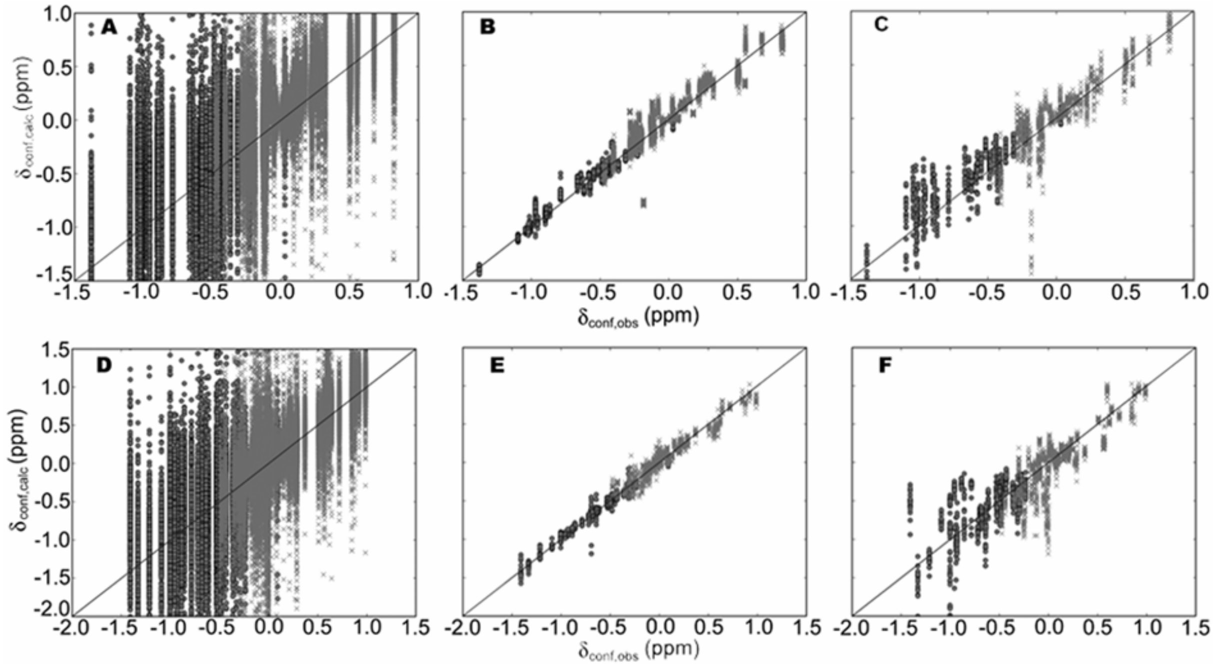


Figure 5.5: Shown are the correlations between observed (horizontal axes) versus back-calculated (vertical) conformational proton chemical shifts for HIV-1 RNA (A-C) and SIV RNA (D-F). The correlations are displayed for starting ensemble (A: $R_{corr} = 0.01$, $RMSD_{chemshift} = 4.0 \pm 56$ ppm; D: $R_{corr} = 0.12$, $RMSD_{chemshift} = 1.6 \pm 3.6$ ppm), final accepted ensemble (B: $R_{corr} = 0.97$, $RMSD_{chemshift} = 0.118 \pm 0.003$ppm; E: $R_{corr}=0.99$, $RMSD_{chemshift} = 0.079 \pm 0.009$ppm), and the deposited PDB ensemble (C: $R_{corr} = 0.92$, $RMSD_{chemshift} = 0.19 \pm 0.02$ ppm; F: $R_{corr}=0.85$, $RMSD_{chemshift} = 0.27 \pm 0.02$ ppm). Note that for the final accepted set of the SIV RNA the chemical shift correlations are displayed of the final accepted structures obtained by backbone regularization with published distance restraints, but without application of RDC restraints. In the correlation plots, filled circles represent aromatic chemical shifts and the crosses sugar chemical shifts.

Interestingly, in the correlation plot of the final accepted chemical shift structures of HIV RNA (B) one chemical shift (G14 H1') lies somewhat off the diagonal. In the pdb ensemble, $\Delta\delta_{conf}$(= $\delta_{conf,obs}$ - $\delta_{conf,calc}$) of G14 H1' varies between -1.51 to 0.10 ppm. In the final accepted structures $\Delta\delta_{conf}$ is smaller and stable around value of -0.55 ppm, which is still relatively large. Analysis shows that the variation $\Delta\delta_{conf}$ in the pdb ensemble is largely caused by variation in buckle (variance 6°), propeller twist (variance 10°), and opening (variance 12°) of the A10-A13 non-canonical basepair. This basepair was kept fixed during structure calculation in a conformation defined by the average the helix parameter values of the pdb ensemble. For the SIV_RNA ensemble of accepted chemical structures (Figure 5E), the values are unbiased for all chemical shifts and show little variation, except for A11 H2 ($\Delta\delta_{conf}$: -0.41 ppm to 0.22 ppm at $\delta_{conf,\ obs}$: -0.69ppm). This relatively wide range is caused by the variation in the conformation of the non-canonical G12-A23. This basepair closes the loop, lies directly above A11 and was left free to adjust within restraint bounds during backbone regularization. The large $\Delta\delta_{conf}$ for A11 H8 in the ensemble of pdb structures (at $\delta_{conf,\ obs}$ -1.41 ppm, Figure 5.5F), has completely disappeared in the accepted chemical shift structures**.**

Figure 5.5 shows the correlations of $\delta_{conf,obs}$ with $\delta_{conf,calc}$ for the starting, final-accepted, and pdb-deposited ensembles of the HIV and SIV RNA stems. The final accepted structures (Figures 5.5 B and E) show a significantly improved correlation over that of the pdb-deposited structures (Figures 5.5 C and F).

For HIV1_RNA and SIV RNA, the RMSD of the helix parameters of the ensemble of final accepted structures ranges from 0.6-1.2$\sigma_{helixpar}$, indicating that the structures are determined with high precision. The final accepted structures closer resemble a standard A-helix than the pdb structures (Tables S5.16a/b vs. S5.19a/b). For HIV RNA, most striking is the absence of the unusual positive signs of propeller-twist of all basepairs and too large helical rise of basepairs G2:C21, C3:G20, and C4:G19 (Table S5.16a vs S5.19a). Leaving end-effects aside (basepairs G1:C22/G2:C21, C9:G12/A10:A11), deviations from standard A-helix values larger than $2.1\sigma_{\text{helixpar}}$ have disappeared in the final accepted structures, except for C8:G15 (deviation $2.2\text{-}2.6\sigma_{\text{helixpar}}$). The latter is most likely transference of end-effects from the closing A:A/C:G basepairs of the tetra-loop. SIV_RNA shows no deviations larger than $2.2\sigma_{\text{helixpar}}$, leaving end-effects aside (Table S5.16b/S5.19b). In conclusion, the average helix parameters of the final accepted structure of the two RNAs indeed fall within the allowed ranges for A-helices with Watson-Crick basepairs, and the structures satisfy the experimental NOE and RDC restraints.

5.5 Concluding Remarks

We presented an efficient method for determining the nucleic acid helix structure with high precision and accuracy that uses exclusively proton chemical shifts as experimental restraints. We have used simulated data to establish in a defined manner the effect of errors and data limitations of the method. We further demonstrated via two deposited RNA structures that also experimental proton chemical shifts suffice to derive RNA helix structures with high precision and accuracy.

A few concluding remarks on the possibilities and limitations of CHEOPS appear in order. Structures derived with CHEOPS are fully based on chemical shifts as no use is made of sequence homology like in chemical shift based structure programs like CS ROSETTA(22), CS23D (24), and CHESHIRE (21). The method works for helices with or canonical, and most likely with, non-canonical basepairs and is connected with a restrained molecular dynamics protocol. The restrained molecular dynamics allows for regularization of the backbone as explicitly shown here, but also for full structure calculation of tertiary structures. We wanted to demonstrate here that unique high precision is achieved for helix structures derived from chemical shifts with the proposed method. We implemented the SVD procedure but the method does allow for a grid search of the helix parameter space to find optimal conformation, e.g. to establish the conformation of a non-canonical basepair. It is also possible to keep certain helix parameters fixed. To employ the helix parameter description within the context of single strand is an obvious extension of the method. The calls to 3DNA allow for single-stranded RNA/DNA molecules. A further and possibly even more important point is that loop structures are commonly dynamic to some degree. To account for dynamics in structures solely based on chemical shifts or on any NMR restraints would require additional method development, e.g. inclusion of ensemble averaging. As for all NMR parameters, the chemical shift is also susceptible to dynamics and this opens new opportunities for assessing dynamics.

Acknowledgment: The 6th framework program of the EU, project FSG-V-RNA, is acknowledged for funding.

5.6 References

1. Kuszewski JJ, Schwieters CD, Clore GM. 2001. Improving the Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base–Base Positional Interactions. *J Am Chem Soc* 123:3903-18
2. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387
3. Pardi A, Hare DR, Wang C. 1988. Determination of DNA structures by NMR and distance geometry techniques: A computer simulation. *Proc Natl Acad Sci USA* 85:8785-9
4. Gronenborn AM, Clore GM. 1989. Analysis of the Relative Contributions of the Nuclear Overhauser Interproton Distance Restraints and the Empirical Energy Function in the Calculation of Oligonucleotide Structures Using Restrained Molecular Dynamics. *Biochemistry* 28:5978-84
5. Allain FHT, Varani G. 1997. How Accurately and Precisely Can RNA Structure be Determined by NMR? *J Mol Biol* 267:338-51
6. Varani G, Aboul-ela F, Allain FHT. 1996. NMR investigation of RNA structure. *Prog NMR Spectrosc* 29 51-127
7. Borgias BA, James TL. 1990. MARDIGRAS - A Procedure for Matrix Analysis of Relaxation for Discerning Geometry of an Aqueous Structure. *J Magn Reson* 87:475-87
8. Keepers JW, James TL. 1984. A Theoretical Study of Distance Determinations from NMR. Two-Dimensional Nuclear Overhauser Effect Spectra. *J Magn Reson* 57:404-26

9. Shajani S, Varani G. 2007. NMR Studies of Dynamics in RNA and DNA by ^{13}C Relaxation. *Biopolymers* 86:348-59
10. Bailor MH, Musselman C, Hansen AL, Gulati K, Patel DJ, Al-Hashimi HM. 2007. Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat. Protoc.* 2:1536-46
11. Mollova ET, Hansen MR, Pardi A. 2000. Global Structure of RNA Determined with Residual Dipolar Couplings. *J Am Chem Soc* 122:11561-2
12. Sibille N, Pardi A, Simorre J, Blackledge M. 2001. Refinement of Local and Long-Range Structural Order in Theophylline-Binding RNA Using ^{13}C-^{1}H Residual Dipolar Couplings and Restrained Molecular Dynamics. *J Am Chem Soc* 123:12135-46
13. Musselman C, Pitt SW, Gulati K, Foster LL, Andricioaei I, Al-Hashimi HM. 2006. Impact of static and dynamic A-form heterogeneity on the determination of RNA global structure dynamics using NMR residual dipolar couplings. *J Biomol NMR* 36:235-49
14. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37
15. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, et al. 1989. Definitions and Nomenclature of Nucleic Acid Structure Parameters. *J Mol Biol* 205:787-91
16. Tolbert BS, Miyazaki Y, Barton S, Kinde B, Starck P, et al. 2010. Major Groove Width Variations in RNA Structures determined by NMR and Impact of ^{13}C Residual Chemical Shift Anisotropy and ^{1}H-^{13}C Residual Dipolar Coupling on Refinement. *J Biomol NMR* 47:205-19
17. Mielke SP, Krishnan VV. 2009. Characterization of protein secondary structure from NMR chemical shifts. *Prog NMR Spectrosc* 54:141-65
18. Grzesiek S, Sass HJ. 2009. From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr.Op.Struct.Biol.* 19:585-95
19. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. 2001. Prediction of Proton Chemical Shifts in RNA. Their use in Structure refinement and validation. *J Biomol NMR* 21:11-29
20. Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* 10:337-50
21. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. 2007. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615-20
22. Shen Y, Lange O, Delagio F, Rossi P, Aramini JM, et al. 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685-90
23. Shen Y, Vernon R, Baker D, Bax A. 2009. De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* 43:63-78

24. Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G. 2008. CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496-W502

25. Cornilescu G, Delagio F, Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289-302

26. Lam SL, Chi LM. 2010. Use of chemical shifts for structural studies of nucleic acids. *Prog NMR Spectrosc* 56:289-310

27. Dejaegere A, Bryce RA, Case DA. 1999. An empirical analysis of proton chemical shifts in nucleic acids. In *Modeling NMR Chemical Shifts. Gaining Insight into Structure and Environment,* ed. JC Facelli, ACd Dios:194-206. Washington DC: American Chemical Society. Number of 194-206 pp.

28. Altona C, Faber DH, Hoekzema AJAW. 2000. Double-helical DNA ^{1}H chemical shifts: an accurate and balanced predictive empirical scheme. *Magn Reson Chem* 38:95-107

29. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-21

30. Ribas-Prado R, Giessner-Prettre C. 1981. Parameters for the calculation of the ring current and atomic magnetic anisotropy contributions to magnetic shielding constants: Nucleic acid bases and intercalating agents. *J.Mol.Struct. THEOCHEM* 76:81-92

31. Giessner-Prettre C, Pullman B. 1987. Quantum mechanical calculations of NMR chemical shifts in nucleic acids. *Quart. Rev. Biophys.* 20:113-72

32. Gelbin A, Schneider B, Clowney L, Hsieh SH, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Am Chem Soc* 118:519-29

33. Arnott S, Hukins DWL. 1973. Refinement of the Structure of B-DNA and Implications for the Analysis of X-ray Diffraction Data from Fibers of Biopolymers. *J Mol Biol* 81:93-105

34. Brunger AT. 1992. *X-PLOR version 3.1 A system for X-ray Crystallography and NMR*. New Haven: Yale University Press

35. Case DA. 1995. Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* 6:341-6

36. Olson WK, Esguerra M, Xin Y, Lu X. 2009. New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures. *Methods* 47:177-86

37. Staple DW, Butcher SE. 2003. Solution structure of the HIV-1 frameshift inducing stem-loop RNA. *Nucleic Acids Res* 31:4326-31

38. Marcheschi RJ, Staple DW, Butcher SE. 2007. Programmed Ribosomal Frameshifting in SIV Is Induced by a Highly Structured RNA Stem–Loop. *J Mol Biol* 373:652-63

5.7 Supplementary Material

Relation between chemical shift and helix parameters

The formal basis for the proposed method for nucleic acid helix structure calculation from chemical shifts is presented below. An important aspect is that conformational chemical shifts are expressed in terms of helix parameters rather than atomic coordinates (1-6). They provide a quantitative framework (5) for describing the conformation of nucleic acids, and thus for understanding the, sometimes subtle, variations in their three-dimensional structures. Also, atomic models can be reconstructed from given helix parameters (4). The reduced space (12 per base pair) spanned by the helix parameters allows one to focus on the essential conformational features rather than having to consider all (partly irrelevant) degrees of freedom, as would be the case with when nucleic acid conformation is described in terms of conventional atomic coordinates or torsion angles. In the nucleic acid structure derivation from proton chemical shifts we present here, we use the program 3DNA and its up-to-date helix parameter description and its atomic-model rebuilding facilities (4).

The 12 helix-parameters can be divided into 6 ‘basepair’ parameters and 6 ‘basepair-step’ parameters. The 6 ‘basepair’ parameters define the conformation of the isolated basepair and can be further grouped into 3 translational parameters (shear, stretch, and stagger) and 3 rotational parameters (buckle, propeller twist, and opening). As for the ‘basepair’ parameters, the 6 ‘basepair-step’ parameters can be grouped into 3 translational parameters (X-displacement, Y-displacement, helical rise or h-rise) and 3 rotational parameters (inclination, tip, helical twist or h-twist).

When considering a proton *k* in a helix consisting of *n* base pairs, its conformational chemical shift ($\delta_{conf,\,k}$) can be expressed as follows as a function of the complete set of helix parameters that describe the helix structure:

$$\delta_{conf,k} = g\left(S_{x,1},S_{y,1},S_{z,1},\kappa_1,\omega_1,\sigma_1,dx_1,dy_1,D_{z,1},\eta_1,\theta_1,\Omega_1,\ldots,S_{x,n},S_{y,n},S_{z,n},\kappa_n,\omega_n,\sigma_n,dx_n,dy_n,D_{z,n},\eta_n,\theta_n,\Omega_n\right) \tag{5.7}$$

Here, *g* indicates the function that transforms helix parameters into chemical shifts. The arguments of *g* are the helix parameters. The chemical shift change, $\Delta\delta_{conf,k}$, resulting from a change of one or more helix parameters can be calculated by a Taylor expansion of the function *g*. For small changes, such a series can be truncated to first order leading to equation(5.8).

$$\Delta\delta_{conf,k} \approx \sum_{i=1}^{n}\sum_{p=1}^{12}\left(\frac{\partial\delta_{conf,k}}{\partial H_{i,p}}\right)_{H=H_{i,p}^{o}} \Delta H_{i,p} \tag{5.8}$$

Here, $H_{i,p}$ is the helix parameter *p* (running from 1 to 12) of basepair *i* (running from 1 to n basepairs); $\Delta H_{i,p}$ indicates the change in helix parameter $H_{i,p}$ around its current value $H_{i,p}^{o}$. The partial derivatives in equations (5.8) have to be evaluated numerically. For instance, the partial derivative $\left(\frac{\partial\delta_{conf,k}}{\partial S_{x,i}}\right)_{S_{x,i}^{o}}$ can be evaluated by changing $S_{x,i}$ around its current value, $S_{x,i}^{o}$, while keeping all other helix parameters fixed at their current values and then evaluating the corresponding conformational shifts:

$$\frac{\partial\delta_{conf,k}}{\partial S_{x,i}} = \frac{(\delta_{conf,k})_{+} - (\delta_{conf,k})_{-}}{(S_{x,i})_{+} - (S_{x,i})_{-}} = \frac{\Delta\delta_{k}}{\Delta S_{x,i}} \tag{5.9}$$

with

$$(S_{x,i})_{+} = S_{x,i} + \frac{\Delta S_{x,i}}{2} \quad (S_{x,i})_{-} = S_{x,i} - \frac{\Delta S_{x,i}}{2} \tag{5.10}$$

In the last term on the right-hand side above and in the following we have dropped the subscript *conf*, so that $\delta_{conf,k}$ becomes δ_k and the superscript *o* so that $S_{x,i}^{o}$ becomes $S_{x,i}$. The equations describing the change in proton chemical shifts variation can now be written in matrix form:

$$\begin{pmatrix} \Delta\delta_{\text{Base 1,H1'}} \\ \vdots \\ \Delta\delta_{\text{Base n,H6}} \\ \Delta\delta_{\text{Base n+1,H1'}} \\ \vdots \\ \Delta\delta_{\text{Base 2·n,H6}} \end{pmatrix} = \begin{pmatrix} \frac{\partial\delta_{\text{Base 1,H1'}}}{\partial S_{x,\text{bp.1}}} & \cdots & \frac{\partial\delta_{\text{Base 1,H1'}}}{\partial\Omega_{\text{bp.n}}} \\ \vdots & \cdots & \vdots \\ \frac{\partial\delta_{\text{Base n,H6}}}{\partial S_{x,\text{bp.1}}} & \cdots & \frac{\partial\delta_{\text{Base n,H6}}}{\partial\Omega_{\text{bp.n}}} \\ \frac{\partial\delta_{\text{Base n+1,H1'}}}{\partial S_{x,\text{bp.1}}} & \cdots & \frac{\partial\delta_{\text{Base n+1,H1'}}}{\partial\Omega_{\text{bp.n}}} \\ \vdots & \cdots & \vdots \\ \frac{\partial\delta_{\text{Base 2·n,H6}}}{\partial S_{x,\text{bp.1}}} & \cdots & \frac{\partial\delta_{\text{Base 2·n,H6}}}{\partial\Omega_{\text{bp.n}}} \end{pmatrix} \bullet \begin{pmatrix} \Delta S_{x,\text{bp.1}} \\ \vdots \\ \Delta\Omega_{bp.n} \end{pmatrix} \tag{5.11}$$

Or, in compact notation:

$$\Delta\mathbf{C} = \mathbf{G}\cdot\Delta\mathbf{H} \tag{5.12}$$

Equation (5.12) provides an expression to calculate the change in proton chemical shifts upon a (small) change in the initial helix structure. By inverting equation(5.12), the change in helix parameters resulting from a change in chemical shift can be calculated. As shown in the results section, within the helix conformational space helix parameters and chemical shifts are related in a smooth, often nearly linear fashion. If linear relations apply between helix parameters and chemical shifts, a matching structure can be calculated directly from a given set of experimental chemical shifts (δ_{exp}) and some starting helix structure using equation (5.12). The effects of non-linearity's can be accounted for by iteratively repeating the procedure illustrated above. This forms the basis for the method for the calculation of a nucleic acid helix structure from chemical shifts presented here.

Technical details on SVD calculation protocol

First a brief overview and main equations of the protocol is given (A), after which a number important assumptions regarding the Singular value Decomposition (SVD) is discussed (B).

(A) *Calculation protocol, main equations.* An overview of the calculation protocol can be found in Scheme S1 (discussed in the next subsection). The linear relation in equation (5.12) holds for small enough changes in helix parameters. In equations (5.11) and (5.12), the helix parameters in **ΔH** and in **G** are expressed in units of standard deviations of RNA A-helix space given in Table S5.1. The aim is to solve equation (5.12), given a list of chemical shifts (**ΔC**). This requires knowledge of the G-matrix. The G-matrix elements are numerically evaluated as described in the main text; NUCHEMICS(7) is thereby employed to calculate the change in chemical shift of the (non-exchangeable) protons upon a small change in the current structure. Given the G-matrix, equation (5.12) can now be solved by calculating the Moore-Penrose inverse of the G-matrix and determining the vector **ΔH** from(8),

$$\Delta\mathbf{H} = \mathbf{G}^{-1} \bullet \Delta\mathbf{C} \tag{5.13}$$

The required inverse of the G-matrix in equation (5.13) can be obtained via singular value decomposition (SVD) as discussed below. The new helix parameters, H_{new}, are subsequently calculated from ΔH,

$$\mathbf{H}_{new} = \mathbf{H}_{\mathbf{0}} + \kappa \cdot \Delta\mathbf{H} \tag{5.14}$$

Here, $\mathbf{H_0}$ is the vector that contains the previous helix parameters. The factor κ is a user chosen parameter (default value=0.50). It can be used to reduce the size of the solution vector **ΔH**. Keeping the factor κ small prevents too large changes in **ΔH**. This is needed, because proton chemical shifts and helix parameters are often close to but not completely linearly related. Consequently, an iterative procedure is followed to find the final solution. A small value for the factor κ assures that, in the iterative procedure, subsequent solution vectors **ΔH** follow the potential curvature of the helix parameter-chemical shift space. The vector $\mathbf{H}_{\mathbf{new}}$ is used in each step in the iterative procedure to rebuild an atomic model of the new helix using 3DNA(4).

(B) *Singular Value Decomposition, assumptions.* We now consider a number of important assumptions in this calculation protocol. First, we describe the derivation of the inverse of the G-matrix by means of SVD and consider which conditions should hold. Any matrix, here matrix G, for which the number of rows (*M*) is equal or exceeds the number of columns (*N*), can be decomposed into(8):

$$\mathbf{G} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^{\mathbf{T}} \qquad (5.15)$$

Here, **Σ** is a N-diagonal matrix, with the N ordered eigenvalues (w_j) of the G-matrix; **U** is a column ortho-normal *M* x *N* matrix, $\sum_{i=1}^{M} U_{ik} U_{in} = \delta_{kn} \quad \left(1 \le k \le N \vee 1 \le n \le N\right)$; V is an ortho-normal N x N square matrix, $\sum_{j=1}^{N} V_{jk} V_{jn} = \delta_{kn} \quad \left(1 \le k \le N \vee 1 \le n \le N\right)$. Given equation(5.15), the Moore-Penrose inverse of matrix G can be written as:

$$\mathbf{G}^{-1} = \mathbf{V} \cdot \left(\Sigma^{-1}\right)^{\mathbf{T}} \cdot \mathbf{U}^{\mathbf{T}} \qquad (5.16)$$

In brief, equation (5.16) applies when(8) $M \ge N$, that is singular value decomposition of the G-matrix can be performed when the total number of proton chemical shifts (rows, *M*) is equal or exceeds the total number of helix parameters (columns, *N*). Note that per base pair up to ~ 15 chemical shifts of non-exchangeable protons can be measured and that 12 helix parameters describe the conformation of a base pair. The number of chemical shifts (*M*) can thus be equal or larger than the total number of helix parameters (*N*). Consequently, equation (5.16) holds, $\mathbf{G}^{-1}$ can be determined, and **ΔH** calculated via equation (5.5) from the chemical shift vector **ΔC**. In fact, the solution vector **ΔH** is then the least-squares solution(8).

What happens if $M < N$? The SVD approach(8) is then to substitute a zero for each element in the **ΔC** vector up to *N* and for each element of the corresponding rows of the **G** matrix. In this way, the condition $M \ge N$ always holds, $\mathbf{G}^{-1}$ can be determined for any value of *M* and the solution vector **ΔH** calculated. The fact that some rows may now contain only elements with value zero needs special consideration(8). Two questions appear important for the solutions obtained(8). (1) How do round off or other errors and relative sizes of the eigenvalues affect the solutions? (2) How does the relative number of chemical shifts versus helix parameters affect the nature of the solution?

(1) *How do errors (round off or other) and relative sizes of the eigenvalues affect the solutions?* The Moore-Penrose inverse of G is susceptible to round-off errors and other types of errors in the G-matrix(8). Round-off errors occur when eigenvalues w_j in **Σ** become very small, e.g. reach the machine precision. The susceptibility to such errors comes about, because the eigenvalues (w_j) may have large differences. Consequently, the inverse values of (w_j), which make up an important component of the inverse of the matrix **G**, may have wildly different values. As a further consequence, the solution vector **ΔH**, may also have wildly large components and their algebraic cancellation when multiplied by the matrix **G**, may give a very poor approximation to the left-hand vector **ΔC** (see equation (5.12)) (8). The solution to this problem is zeroing small eigenvalues (8). In fact, the solution vector **ΔH** obtained by zeroing the small w_j's and then using equation (5.13) and (5.16) is often more stable, than the SVD solution where the small w_j's are left nonzero (8). There is another reason for editing even additional eigenvalues, like ones large enough so that round-off errors are not a question. Singular value decomposition allows one to identify linear combinations, which do not contribute much to reducing the squared residuals χ^2 of the data set (8),

$$\begin{aligned}\chi^2 &= (\mathbf{G}\Delta\mathbf{H} - \Delta\mathbf{C})^2 = \\ &= (U_{(1)}V_{(1)}^T w_1 + U_{(2)}V_{(2)}^T w_2 + \ldots + U_{(n)}V_{(n)}^T w_n + ..)\Delta\mathbf{H} - \Delta\mathbf{C})^2\end{aligned} \tag{5.17}$$

Here, $U_{(m)}$ and $V_{(m)}$ are columns in matrices U and V, respectively. Zeroing the small eigenvalues w_j does not significantly increase χ^2. Furthermore, the squared probable error (variance) of the coefficients of the solution vector **ΔH** is given by (8),

$$\sigma^2(\Delta\mathbf{H}) = \frac{V_{(1)}^2}{w_1^2} + \frac{V_{(2)}^2}{w_2^2} + ..\frac{V_{(n)}^2}{w_n^2} + ..; \tag{5.18}$$

Evidently, zeroing small eigenvalues w_j can reduce the probable error on the coefficients of the solution vector quite significantly, while as discussed above it increases the minimum in χ^2 only marginally(8). Thus, by zeroing small eigenvalues, more stable solutions are obtained, i.e. solutions that are less sensitive to random errors in **ΔC**. Therefore, it is recommended to zero too small eigenvalues and their reciprocals(8).

The criterion to establish whether a singular value is too small was operationally set as follows: the maximum eigenvalue max$\{w_j\}$, was divided by 100 and multiplied by a threshold factor λ (default value is 0.10); if the eigenvalues, w_j are smaller than this value, they are substituted by zero:

$$w_j < \lambda \cdot \frac{\max\{w_j\}}{100} \Rightarrow (w_j) = 0 \vee \left(\frac{1}{w_j}\right) = 0 \tag{5.19}$$

(1) *How does the number of chemical shifts (M') relative to the number of helix parameters (N) affect the solution?* If *M'* ≥ *N*, the system is overdetermined and one obtains the least-squares solution(8). If *M'* < *N*, the system is underdetermined and the solution (**ΔH**) is in general not unique. It can be shown(8) that SVD selects a defined solution. This is because the elements in the vector **ΔC** between *M'* and *N* and the corresponding rows in G are zeroed. The solution obtained is then the least-squares solution in reduced space (*M'* dimensional space), while the coefficients of the solution vector **ΔH** originating from the space between *M'* and *N* are zero. As a result, the solution vector is the one closest to the starting point. Thus, the SVD solution vector depends in this case on the starting point (8). Consequently, when a variety of different starting points is taken, a broad set of solutions is obtained when the system is underdetermined. In contrast, when the system is overdetermined, the solutions do not depend on the starting point. One single solution is then obtained, even when starting from a variety of different starting points. In the main text we describe, how we make use of this property of SVD solutions, i.e. we employ a wide variation in starting points for the structure calculations that cover the allowed RNA A-helix conformational space. This allows us to determine, whether or not or how strongly, a decreasing number of chemical shifts affect the precision of the solutions **ΔH**. Finally, it is important to note that when the number of measured chemical shifts (*M'*) is less than 12 per base pair (*N*), the system appears underdetermined, but in fact, it may not be. The conformational space of RNA/DNA A-helices is restricted by the chemical bonds that link the base pairs and residues, thereby

effectively reducing the degrees of freedom (dof) from 12 to a lower number and we make use of this to select solutions compatible with A-helix context (see main text).

A more extensive discussion on SVD matters can be found in the book, Numerical Recipes in C by Press et al.(8).

Flowchart of CHEOPS

A flowchart of the protocol for the derivation of helical nucleic acid structures from chemical shifts is presented in Scheme S1. Briefly, in its present form the protocol requires as input a set of (experimental) proton chemical shifts and a starting structure with corresponding helix parameters. Equation (5.12) is then solved to obtain a new set of helix parameters and corresponding full-atom output structure. This output structure acts as a starting structure in a new pass though the scheme. This process is repeated until the proton chemical shifts are consistent with the experimental input chemical shifts within a user-specified tolerance. The protocol can further be repeated with different starting structures to generate an ensemble of output structures reflecting solution ranges. A step-by-step description of the calculation protocol is given below.

1) Template PDB file and derivation of helix parameters

A starting helix model structure is read in and analyzed by calling on 3DNA (4) to generate the corresponding set of helix parameters denoted as $\mathbf{H_o}$ (In the first pass, the starting helix model is external, while in later passes through the scheme, see below, the starting helix model is the current updated helix structure). Note that atomic models of the starting helix are generated from helix parameters using the rebuilding feature in 3DNA (4). The residue units that were used, are from the fiber-model A-helix as defined by Olson et al (6). These residue units were employed by us, because initially a fiber-model A-helix was built from a user-provided primary sequence using the fiber-model A-helix building feature in 3DNA(4). Other residue units could be used.

2) *ΔC-vector*

The script then calls the program NUCHEMICS(7) to calculate the proton chemical shifts ($\mathbf{C_o}$) of the current 'starting' helix model. In the first pass, a list of chemical shifts (e.g. experimental chemical shifts) is read in ($\mathbf{C_{in}}$; note that in later passes, $\mathbf{C_{in}}$ is already available). Subsequently, the script calculates the difference between the provided set of chemical shifts ($\mathbf{C_{in}}$) (e.g. experimental shifts) and the calculated values ($\mathbf{C_o}$) to create the vector ΔC (= $\mathbf{C_{in}} - \mathbf{C_o}$).

3) *Convergence check*

The program checks whether all elements of ΔC are within the user-defined chemical shift tolerance (ΔC < threshold). If 'No', the scheme is followed further down. If 'Yes', convergence is reached. The current 'starting' structure then becomes the converged SVD structure with current $\mathbf{H_o}$ and $\mathbf{C_o}$ as helix parameters and chemical shifts. This structure is subjected to backbone regularization, as described below, leading to the final converged output structures. An optional H-bond and backbone quality check is sometimes invoked prior to this backbone regularization. In this case and only if the structure does not show any violations in H-bonds and/or holonomic restraints, the converged SVD structure is directly written out as the final converged output structure.

4) *G-matrix*

All partial derivatives in the G-matrix are calculated numerically for the starting helix model using equation (5.12). Each helix parameter is shifted 0.05 σ (see Table S5.1) above/below the starting value $\mathbf{H_o}$ and all proton chemical shifts in the resulting structures are calculated and the gradients constructed.

5) *Solving equation (5.12) via Singular Value Decomposition to obtain new helix parameters*

Equation (5.12) can be solved by calculating the Moore-Penrose inverse of matrix **G**, $\mathbf{G^{-1}}$, which can be obtained via singular value decomposition (SVD)(8). A solution vector **ΔH** is obtained by multiplying the vector **ΔC** with $\mathbf{G^{-1}}$**.** It is important to note that the SVD solution vector **ΔH** is quite sensitive to small deviations (or errors) in **G** (and thus in $\mathbf{G^{-1}}$ and in **ΔC)**. To overcome these problems, we employed the standard solution of zeroing small eigenvalues (see also Supplementary Material)(8). By zeroing small eigenvalues, more stable solutions are obtained (i.e. solutions that are less sensitive to random errors in **G** and **ΔC)**. The threshold for zeroing small eigenvalues is set by means of a factor λ, i.e. when an eigenvalue of the G-matrix is smaller than λ times the largest eigenvalue, it is set to zero (and its reciprocal) and not considered. An optimal value for λ is found to be ca 0.1 % for the present calculations.

6) *Generate new set of helix parameters*

The new helix parameters $\mathbf{H_{new}}$ are obtained by adding the solution vector **ΔH**, scaled by a factor κ, to the vector containing the set of original helix parameters, $\mathbf{H_{new}} = \mathbf{H_o} + \kappa\,\mathbf{\Delta H}$. Keeping the factor κ smaller than 1 (default value=0.5) ensures that consecutive solution steps in the iterative procedure follow the potential curvature of the helix parameter-chemical shift space better.

7) *Building of a full-atom model of the new helix*

The rebuilding feature of 3DNA(4) is used to build a new full-atom structure model, defined in Cartesian space, based on the new set of helix parameters ($\mathbf{H}_{\mathbf{new}}$). This conversion is carried out, because NUCHEMICS requires input structures that are defined in Cartesian space.

Steps 1 to 7 are repeated iteratively until all calculated chemical shifts match the experimental chemical shifts within a user-set tolerance limit (default 0.05 ppm), or until the maximum number of calculation cycles (default 150) are reached.

The calculation protocol was written in the object-oriented programming language Python v2.4.3 and uses the numerical package numpy 1.0.4-1. The script runs and was fully tested in a LINUX environment. In this paper, 3DNA(4) is used for analysis and rebuilding of nucleic acid structures (helices here). However, the script works equally well with 3DNA v2.0(9). The program NUCHEMICS (7; 10) was employed to calculate the ^{1}H chemical shifts. It makes use of the parameter sets proposed by Giessner-Prettre (11; 12) (GP set) to define the ring current strengths and magnetic anisotropies or the ring current parameter set proposed by Case (13) (DC set).

Investigation of the sensitivity of nucleic acid helix structure to changes in chemical shift, and vice versa. Chemical shifts are highly sensitive probes of structural changes. To put this notion on a quantitative footing G-matrices were calculated of RNA A-helices whose conformations fall within the allowed A-helix conformational space. The elements of the G-matrices are the partial derivatives of the proton chemical shift with respect to the helix parameters. For the calculations an RNA A-helix of 6 base pairs (5'ACAGCU-3':5'-AGCUGU-3') was used. Its conformation was built based on average A-helix parameter values in Table S5.1. The following main results are obtained.

a) The sensitivity of the chemical shift to changes in helix parameters can be represented by the summed absolute values of the row elements of the G-matrix, as follows from:

$$\left|\Delta C_n\right| = \sum_{j=1}^{m}\left|G_{nj}\Delta H_j\right| \approx \sum_{j=1}^{m}\left|G_{nj}\right| \approx \sum_{j=res(i-1)}^{res(i+1)}\left|G_{nj}\right| \qquad (5.20)$$

Here, ΔC_n is the change in chemical shift of proton n, ΔH_j is the change in helix parameter j, G_{nj} denotes the element nj of the G-matrix, m denotes the total number of helix parameters, i.e. m = 6 (base pairs) * 12 (helix parameters per base pair), and n represents a proton chemical shift. The one-to-last equality holds when i) G_{nj} is expressed in units of ppm/σ_{helixpar} (σ_{helixpar} is given in Table S5.1) and ii) ΔH_j equals 1 for all helix parameters (expressed in units of σ_{helixpar}). The last equality follows from the approximately block-diagonal shape of the G-matrix. We find that the sensitivity of chemical shifts to structural changes follows the trend: H5≈H2 > H6≈H2'≈H8 > H1'≈H3' > H5'≈H4' > H5". The chemical shift of H4' and also H5" change relatively little (~ 0.2 ppm/σ_{helixpar}), whereas the chemical shift of H5 and H2 protons change by more than 2.0 ppm/σ_{helixpar} and 2.7 ppm/σ_{helixpar}, respectively. Note that the experimental error in the chemical shift ($\sigma_{\text{chemshift}}$) is often considered to be of the order of 0.01 to 0.02 ppm, so that even the chemical shifts of H4' and H5" are expected to contribute significantly to the structure definition. A detailed overview of summed absolute G-matrix values of bases C2 and A3 are listed in Table S5.2.

b) The summed absolute values of the elements of the rows of **G-1** represent a qualitative measure for the sensitivity of a helix parameter for chemical shift changes:

$$\left|\Delta H_j\right| = \sum_{n=1}^{p}\left|G_{jn}^{-1}\Delta C_n\right| = \sum_{n=1}^{p}\left|G_{jn}^{-1}\right| \approx \sum_{n=res(i-1)}^{res(i+1)}\left|G_{jn}^{-1}\right| \qquad (5.21)$$

Here, p denotes the total number of chemical shifts, i.e. $p = 6$ (base pairs)* 15 (chemical shifts per base pair). The one-to-last equality holds when ΔC_n equals 1 ppm for all chemical shifts. The last equality follows from the approximately block-diagonal shape of the G-matrix and G^{-1}-matrix. The summed absolute G^{-1} values do provide the relative order. However, they cannot be used to evaluate the change in helix parameter from changes in chemical shift in a quantitative manner. Tables S5.2 (column 4) reports the summed absolute G^{-1} values corresponding to the helix parameters of base A3. We find that the sensitivity to changes in the chemical shift follows the order, *buckle* > *Y-displacement* ≈ *tip* > *propeller twist* > *inclination* ≈ *rise* > *helix-twist* ≈ *X-displacement* > *stagger* > *opening* > *shear* > *stretch*. The three least sensitive parameters - shear, stretch and opening - were kept fixed during the structure calculations at the (standard) helix values as given for instance in Table S5.1.

Generation of helices with a random conformation within allowed A-helix conformational space. To prevent bias in the resulting set of chemical-shift-based structures, the structure calculation protocol (Scheme S1) was repeated employing as ‘input pdb files’ a set of random starting structures that cover the complete allowed A-helix conformational space. That is, each of the helix parameters was randomly varied around its standard A-helix values over the interval $[-x\sigma_i, x\sigma_i]$, where σ_i is the standard deviation of the i^{th} helix parameter (see Table S5.1 in the results section) and x is set to 3, except otherwise stated. Note that with x set to a value of 3, the starting structures cover 99.7% of the allowed helix space. As described above, the atomic models of the starting helices were generated from helix parameters using the rebuilding feature in 3DNA (4) and using residue units from the fiber-model A-helix as defined by Olson et al (6).

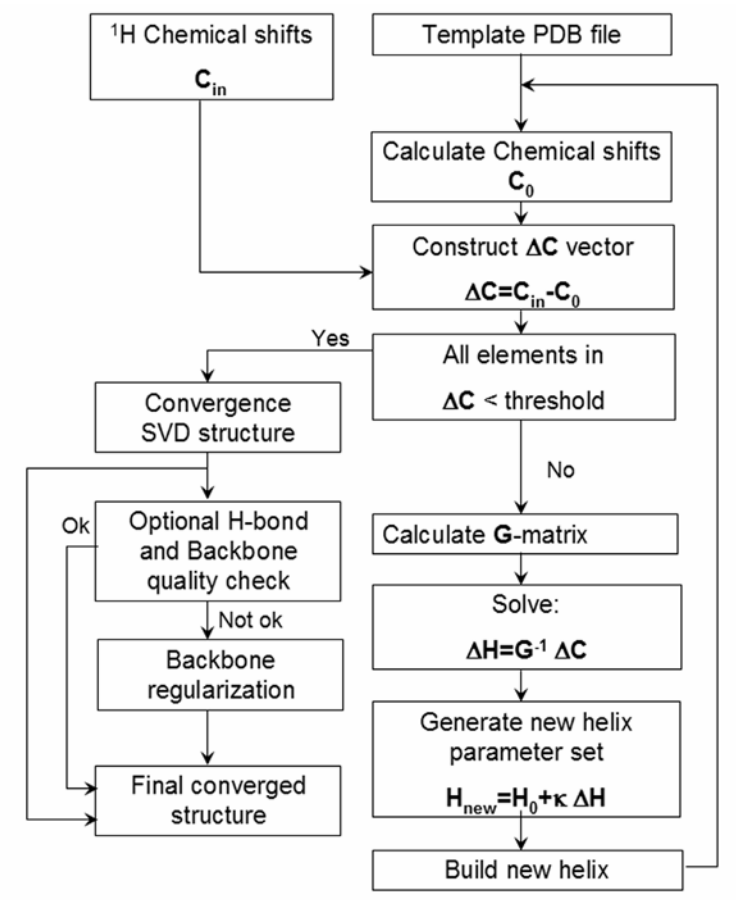
Effect of variation in ring current parameter values. To estimate the effect of uncertainty in parameterization of the ring current and magnetic anisotropy on the chemical shift derived structures the foregoing test calculations were repeated, using the more recent parameter set derived by D. Case(13) (DC set) instead of the above used GP parameter set(11; 12) to calculate chemical shifts in the structure calculation protocol. Note that the DC parameter set incorporates both ring-current and magnetic anisotropy effects into one ring-current parameter. The ‘observed’ chemical shifts were still those calculated from the *target structure* using the GP set. The protocol was run as in the previous section, but with a convergence tolerance of 0.10 and the resulting structures were analyzed in a similar fashion.

To employ the DC parameter set optimally, the reference chemical shift values, δ_{ref}(7; 10) for this parameter set had to be estimated, because they were not made available in the paper by Cromsigt et al.(10). This was done in a fashion, similar to the one employed for the GP parameter set described earlier (7; 10). From the foregoing calculation the best helix, i.e. the helix that showed the lowest chemical shift RMSD with respect to the 'observed' GP chemical shift values was chosen and NUCHEMICS(7; 10) was run on this structure with the DC parameter set to calculate the proton chemical shifts. The reference chemical shifts were adjusted to minimize the chemical shift RMSD of these 'observed' DC chemical shift values with respect to the GP chemical shift values.

The results of the test are given in Table S5.13/S5.14 and Table 5.1 main text.

Backbone regularization. To correct for small errors in the 'geometry/conformation' that may occur in the sugar rings and backbone of the converged SVD structures built by 3DNA, backbone regularization was applied. This regularization was only necessary in the case the calculation of structures was done with an incomplete set of chemical shifts, as occurs in the jackknife testing described below. The backbone regularization is achieved by a Powell minimization of 10000 steps or by a brief restrained Verlet molecular dynamics minimization, run in Xplor v3.851 (14), while keeping the aromatic planes fixed in atomic Cartesian space, so that helix parameters are not affected. Holonomic restraints were included and derived from the 3DNA reference nucleotides (6). Also included were loose standard A-helix NOE distance restraints with the distances taken from Wijmenga et al. (15). The standard A-helix NOE distance restraints consisted of inter-residue distances with upper and lower boundaries of 1.00 Å for H2'$_i$-H1'$_{i+1}$, H2'$_i$-H5'$_{i+1}$, and H2'$_i$-H5$_{i+1}$, upper and lower boundaries of 1.25 Å for H2'$_i$-H6/8$_{i+1}$, and upper boundary of 1.50 Å and lower boundary of 0.5 Å for H1'$_i$-H6/8$_{i+1}$, and H1'$_i$-H5'$_{i+1}$. For the NOE restraints, a soft-square potential was employed with force constant of 45 kcal mole^{-1} (Å)$^{-2}$. Dihedral angles in the backbone and sugar rings as well as the dihedral angle χ were loosely restrained to standard A-helix values as given by Gelbin et al. (16). For the dihedral restraints a harmonic potential was used with force constant of 150 kcal mole^{-1} (°)$^{-2}$. The dihedral angles in the (N-puckered) sugar were restrained within ± 10°, the χ dihedral angles within ± 40°, and the backbone dihedral angles within ± 30°. As a last step, all structures that still contain clashes, according to NUCHECK(17), were removed from the set of resulting structures.

Scheme S5.1 Flowchart of the protocol for the derivation of helical nucleic acid structures from chemical shifts.



The template PDB file is read in, and chemical shifts are calculated from this structure. The calculated chemical shifts are subtracted from the restraint chemical shifts, thereby composing vector **ΔC**. Subsequently, convergence is judged, i.e. are all chemical shifts within a user-set threshold limit? If ‘Yes’, the converged SVD structure is written out, its backbone may be regularized when required and subsequently the final converged structure is obtained. If ‘No’, the scheme followed further down. In addition, the **G**-matrix is calculated as described in the text. Subsequently, the **ΔH** vector is solved and a new set of helix parameters calculated $\mathbf{H}_{\mathbf{new}}$. The **ΔH** vector is scaled by κ in this calculation, to prevent too large changes and thereby restraining the solution to the regime, where the approximation of a linear relationship between chemical shifts and helix parameters holds. From the new helix parameter set, H_{new}, a new atomic model is built, and the calculation enters its new cycle.

Additional figures and tables

Figure S5.1. Additional plots of relationships between H5 proton chemical shift and helix parameters like shown in Figure 5.1 in the main text.

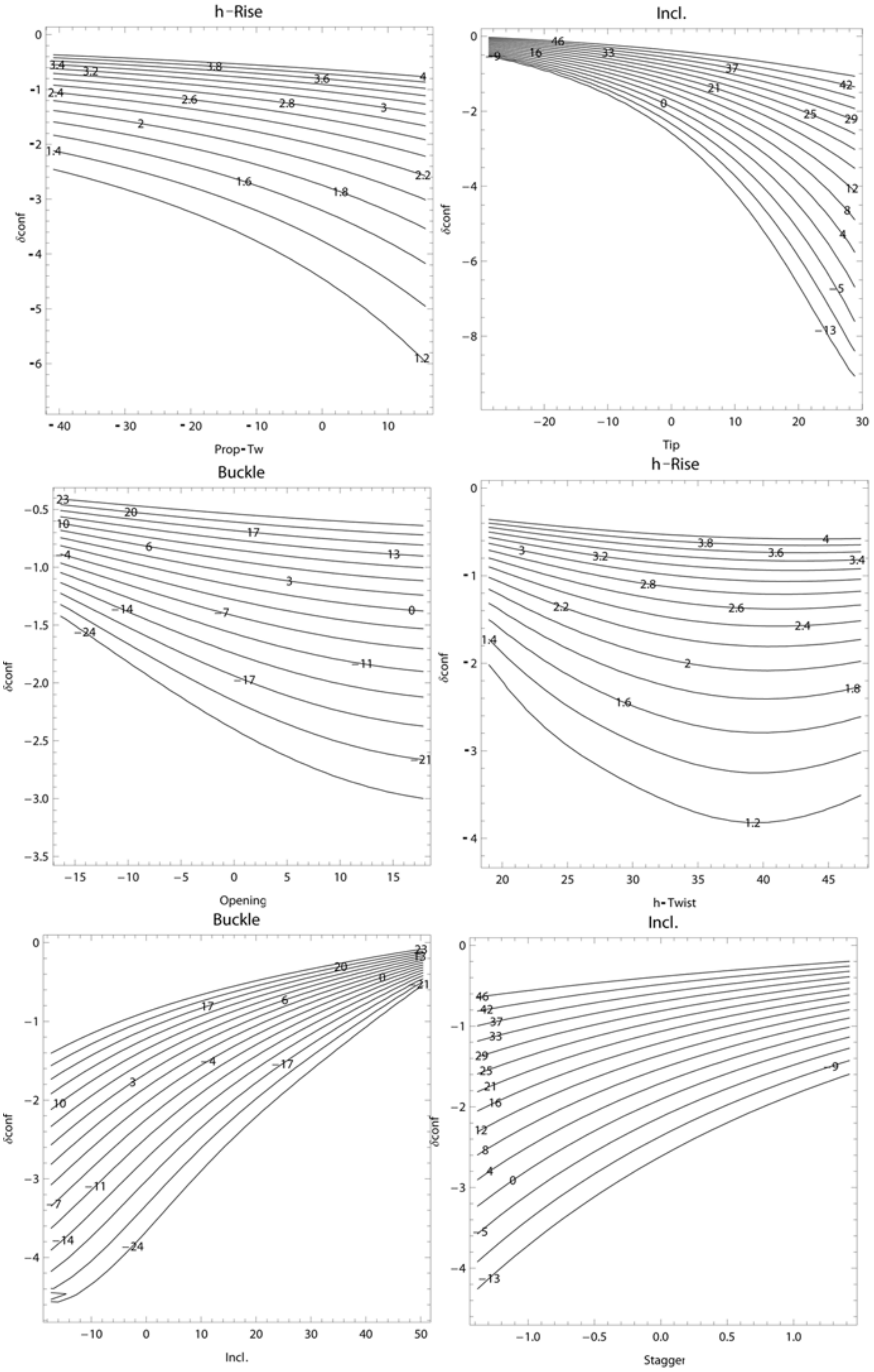
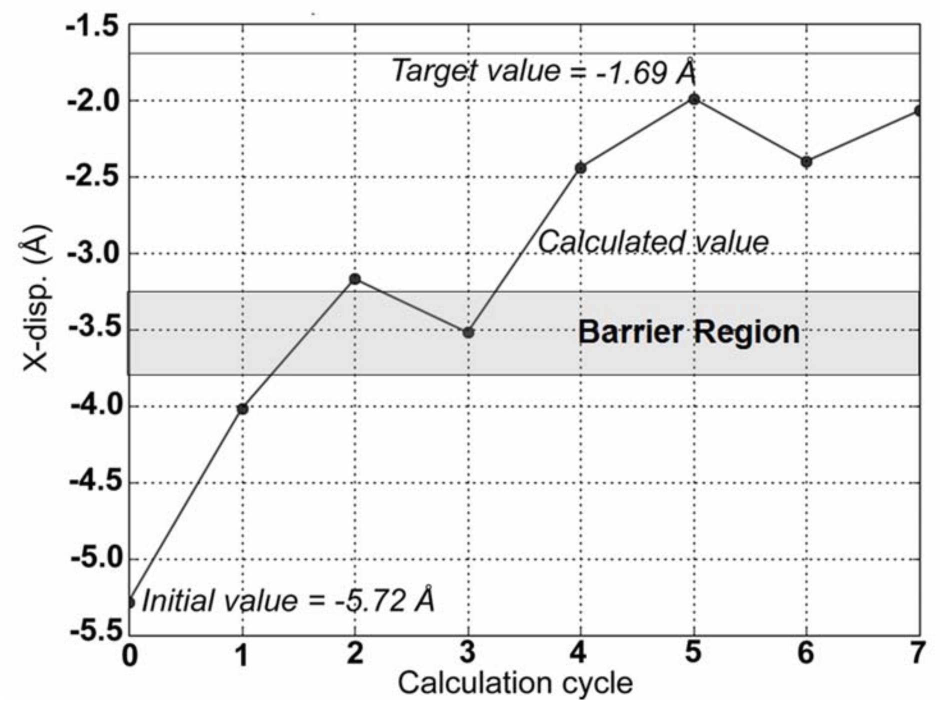


Figure S5.2. Crossing of the minimum (barrier) of X-displacement of a basepair in a test calculation.



In a test calculation, the starting structure had an X-displacement set to -5.72 Å. The target had an X-displacement value of -1.69 Å. This value is located on the other side of the minimum in the curve for the H5 chemical shift versus X-displacement as shown in Figure 5.1e in the main article. The question then is: does one find during the iterative calculation protocol the global solution of –1.69 Å for the X-displacement and thus, is the minimum passed. During the calculation, we followed how the X-displacement of this basepair changes per calculation cycle. In the figure, the horizontal line in the top of the graph indicates the target value; the minimum and small surrounding region of the curve in Figure 5.1e, is indicated here by a gray block called the “Barrier Region”. The dots connected by a line show how the X-displacement of this basepair changes from cycle to cycle. As can be seen, the X-displacement progresses towards the target value and passes the “Barrier Region”. This illustrates that the apparent parabolic behavior of the H5 chemical shift versus X-displacement does not pose a problem in our calculations.

Figure S5.3: Representative histograms of the distributions of helix parameters of basepair 3 in the final accepted structures in the test calculations, where a helix of 6 base pairs with primary sequence 5'ACAGCU3'/3'AGCUGU5' was used as described in the text. Shown are the distributions for helical rise, buckle, and Y-displacement as function of current minus target value on the horizontal axes. The left-hand figures show the deviation in Å or ° units, while the right-hand figures show the deviation in terms of z-scores using the standard deviations of allowed A-helix space (Table S5.1) as normalization factor, as defined in equation (5.22)

$$z = \frac{h_{i,\,finalaccepted} - h_{i,target}}{\sigma_{i,helixpar}} \tag{5.22}$$

Here $h_{i,finalaccepted}$ and $h_{i,\,target}$ are helix parameter i of final accepted and target structure respectively, and $\sigma_{i,\,helixpar}$ is the standard deviation of helix parameter i in Table S5.1). The strength of the dependence of the helix parameters on chemical shift (Table S5.2) is reflected in the width of the distributions in terms of z-scores. The buckle shows the narrowest (z score) distribution with its strong chemical shift dependence.

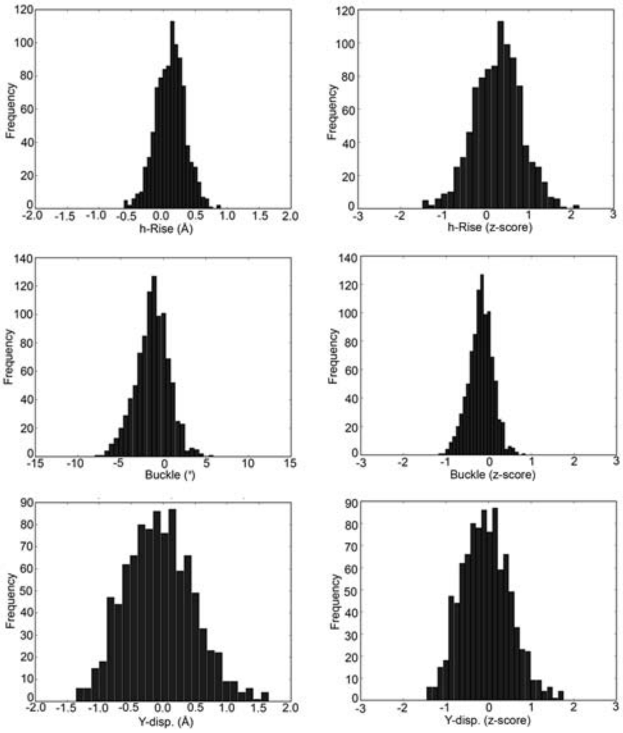
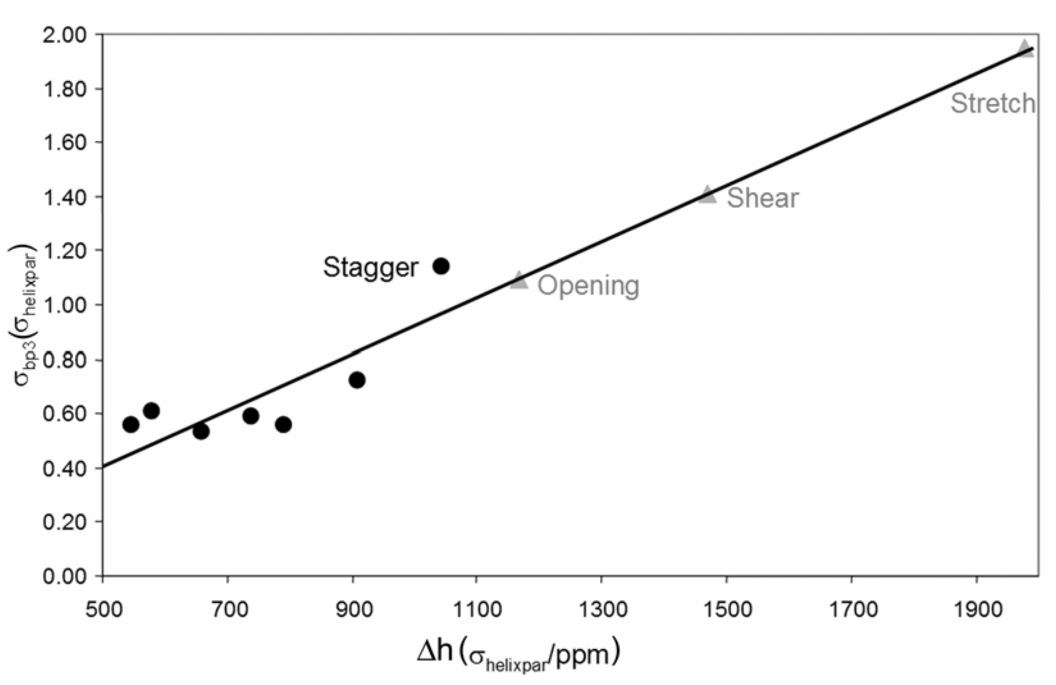
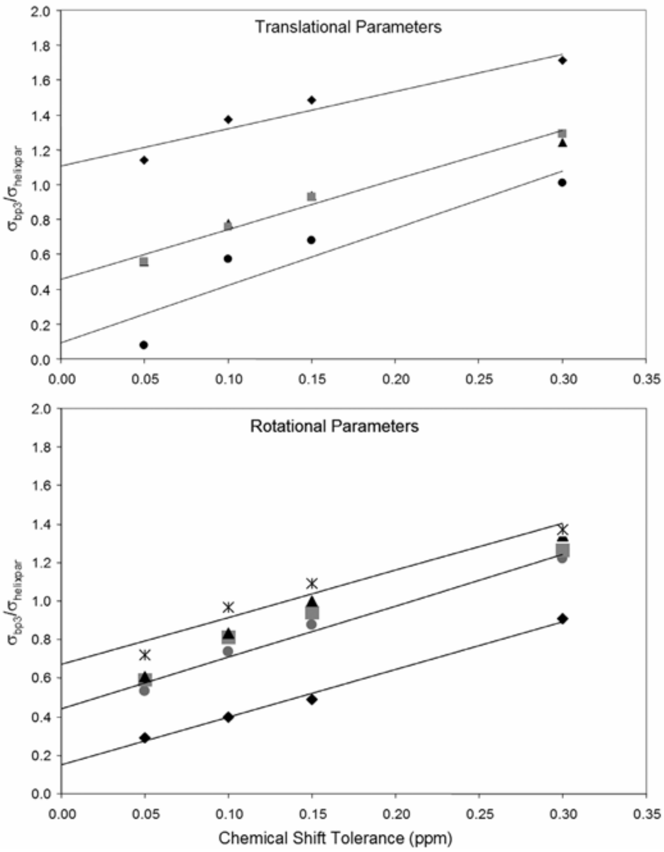


Figure S5.4. Normalized spread in helix parameter, σ_{bp3}, versus Δh values (Table S5.2)



The symbol σ_{bp3} stands for the standard deviation of a helix parameter of basepair 3 in the ensemble of final accepted structures (filled circles); it is normalized to $\sigma_{helixpar}$. The deviation in σ_{bp3} is taken with respect to the corresponding helix parameter in the *target structure*. The symbol Δh is a parameter, which quantifies the sensitivity of a helix parameter to a change in proton chemical shift (here 1 ppm; see Table S5.2). In the structure calculations, the GP-set was employed and the chemical shift tolerance was set to 0.05 ppm, the other parameters are described in the main text. The grey filled triangles represent extrapolated normalized σ_{bp3} values; they are based on the least-squares line, fit to the found values (filled circles).

Figure S5.5. Spread in helix parameter values in final accepted structures as a function of chemical shift tolerance.



The $\sigma_{bp3}/\sigma_{helixpar}$ is the standard deviation of helix parameter of basepair 3 normalized to $\sigma_{helixpar}$ in the ensemble of final accepted structures. In the structure calculations, the GP-set was employed and the chemical shift tolerance was set to 0.05 ppm, the other parameters are described in the main text. The deviation in σ_{bp3} is taken with respect to the corresponding helix parameter in the *target structure*. Top panel, translational parameters: black diamonds – stagger, black circles – X displacement, black triangles – Y displacement, and grey squares - helical rise). Bottom panel, rotational parameters: black diamond – buckle, grey circles – propeller twist, grey squares – inclination, black triangles – tip, and black crosses – helical twist. Note that in the top panel, the black triangles (Y-displacement) and grey squares (helical rise) nearly coincide. In both panels, three least-squares lines are shown through the respective symbols, to visualize the essentially linear increase in the spread in the helix parameters (σ_{bp3}) upon increased chemical shift tolerance.

Figure S5.6. RMSD$_{av2target}$, RMSD$_{pairwise}$, and RMSD$_{chem.shift}$ of accepted helix structures plotted versus chemical shift tolerance

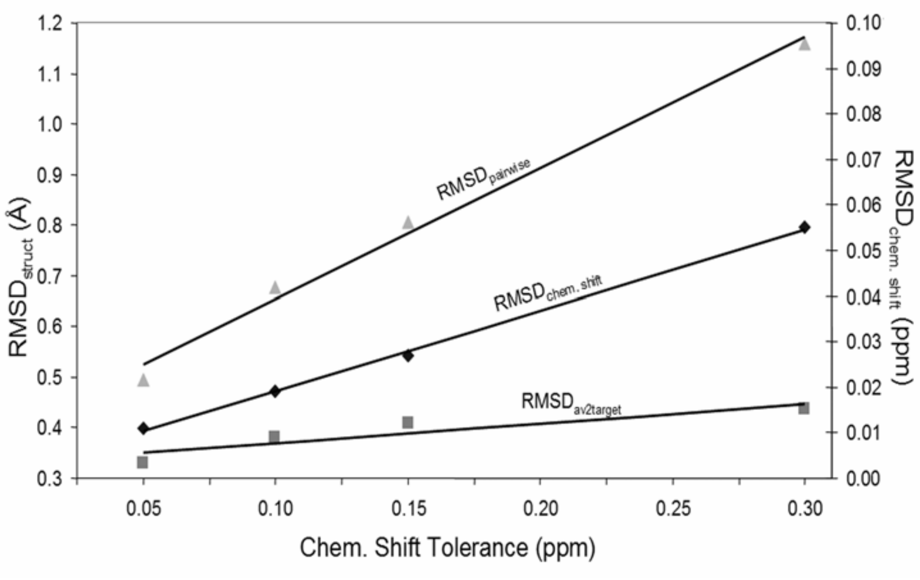
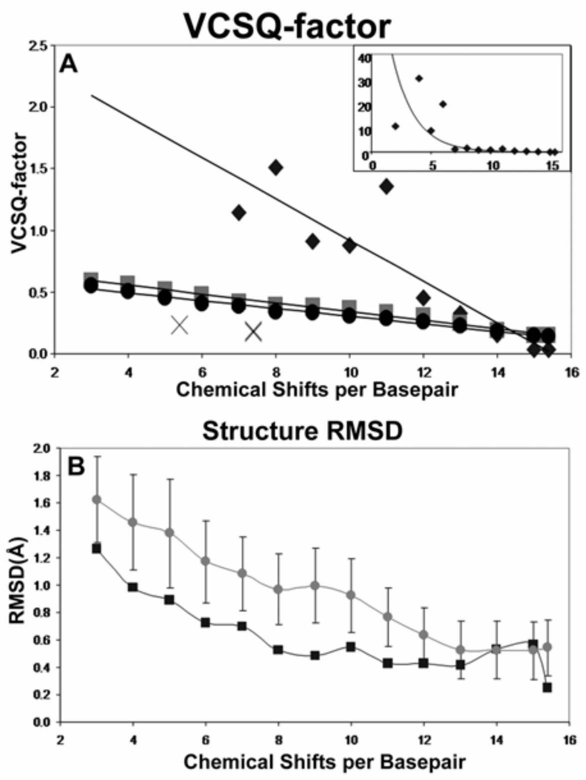


Figure S5.7. The effect of a reduced number of chemical shifts on the precision of the accepted structures. A. VCSQ-factor versus chemical shift per basepair. B. Pairwise RMSD of accepted structures as a measure of precision together with the $\mathrm{RMSD}_{\mathrm{av2target}}$.



A. The VCSQ-factor in relation with the number of chemical shifts per basepair. The large frame shows the effect of backbone regularization and clash removal on the VCSQ-factor; the insert displays the result of the SVD procedure per se. The diamonds show the results of the SVD procedure, the gray squares show the results of the backbone regularization, and the black circles show the results, i.e. after clash removal. Finally, the black crosses show the VCSQ-factor for the results of the special sets, i.e. easily measurable proton shifts. **B**. The RMSDpairwise (grey circles) and RMSDav2target (black squares) of accepted structures (after hydrogen bond check, backbone regularization and check on van der Waals clashes) plotted versus the number of chemical shifts per basepair. The error-bars on RMSDpairwise indicate one standard deviation calculated from the ensemble of accepted structures. For both A and B, a chemical shift tolerance of 0.1 ppm for employed chemical shifts was used as convergence criterion. Further, for restraint sets with less than 7 chemical shifts per basepair, calculations started from 250 random structures instead of the 100 otherwise used; this to compensate for the reduced number of accepted structures.

Table S5.1. Helix parameters of a representative set of RNA A-helices defining A-helical space

Parameter	**A-RNA**$^\$$	**A-DNA**	**B-DNA**
Shear (S_x)	0.00 (0.42) Å	0.01 (0.23) Å	0.00 (0.21) Å
Stretch (S_y)	-0.20 (0.18) Å	-0.18 (0.10) Å	-0.15 (0.12) Å
Stagger (S_z)	0.02 (0.35) Å	0.02 (0.25) Å	0.09 (0.19) Å
Buckle (κ)	-1.01 (6.81) °	-0.10 (7.80) °	0.50 (6.70) °
Prop. Twist (ω)	-12.59 (7.07) °	-11.80 (4.10) °	-11.40 (5.30) °
Opening (σ)	0.74 (4.27) °	0.60 (2.80) °	0.60 (3.10) °
X-displacement (D_x)	-4.30 (1.19) Å	-4.17 (1.22) Å	0.05 (1.28) Å
Y-displacement (D_y)	-0.10 (0.94) Å	0.01 (0.89) Å	0.02 (0.87) Å
h-Rise (D_z)	2.71 (0.41) Å	2.83 (0.36) Å	3.29 (0.21) Å
Inclination (η)	16.55 (8.46) °	14.70 (7.30) °	2.10 (9.2) °
Tip (θ)	0.07 (7.19) °	-0.10 (5.20) °	0.00 (4.30) °
h-Twist (Ω)	33.23 (3.56) °	32.50 (3.80) °	36.5 (6.60) °

$^\$$The set consists of 20 RNA helices, corresponding to 165 base pairs. They were all high-resolution X-ray structures and chosen as described in Materials and Methods. The number indicates the average value of the helix parameter; the standard deviation ($\sigma_{helixpar}$), given within parentheses, forms a measure for the allowed RNA A-helix space. The values for A- and B-DNA are as derived by Olson et al (6) and reported for completeness in column 3 and 4.

Table S5.2. Impact of changes in helix parameters on chemical shifts and vice-versa for all categories of proton shifts

Name[$]	ΔChemical shift[$] (ppm/ σ_{helixpar})		Helix parameter of base pair 3[#1]	Δh[#1] (σ_{helixpar}/ppm)	σ_{bp3} (σ_{helixpar}) after structure derivation[&]
C2-H1’	0.909		Shear (S_x)	1469[#2]	fixed
C2-H2’	1.346		Stretch (S_y)	*1979*	fixed
C2-H3’	0.726		Stagger (S_z)	1045	1.14
C2-H4’	0.283		Buckle (κ)	424	0.29
C2-H5’	0.406		Prop. Twist (ω)	658	0.53
C2-H5”	0.209		Opening (σ)	*1167*	fixed
C2-H5	2.722		X-displacement (D_x)	955	0.42
C2-H6	1.439		Y-displacement (D_Y)	547	0.56
A3-H1’	0.665		h-Rise (D_z)	791	0.56
A3-H2’	1.241		Inclination (η)	739	0.59
A3-H3’	0.675		Tip (θ)	581	0.61
A3-H4’	0.227		h-Twist (Ω)	910	0.72
A3-H5’	0.317				
A3-H5”	0.198				
A3-H2	1.978				
A3-H8	1.102				

[$] Summation of the rows of the G-matrix; as follows from equation (5.20) the number represents (column 2) the sensitivity of chemical shift to change in helix parameter, i.e. the change in chemical shift upon a change in helix parameter equal to one σ_{helixpar}, the standard deviation of a helix parameter in the standard A-helix space (Table S5.1). [#1] The number represents the sensitivity of a helix parameter to a change in chemical shift; it is calculated as the sum of the rows in the Moore-Penrose inverse of the G-matrix as given equation (5.13); as follows from equation(5.16), it is expressed in units of σ_{helixpar} (Table S5.1) per ppm. [#2] Note that the relatively large values can be understood from the fact that 1 ppm corresponds to a relatively large change in chemical shift. [&] σ_{bp3}, the standard deviation of a helix parameter, expressed in units of σ_{helixpar} (Table S5.1), with respect to target in the bundle of structures obtained after structure calculation using chemical shifts restraints (Table 5.1 main text, GP-set, 0.05 ppm N=956).

Table S5.3. Helix parameters of the fiber-model A-helix, i.e. *target structure*.

	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening
A-U	0.01	-0.11	0.06	0.02	-10.52	-2.36
C-G	0.14	-0.18	0.07	-0.19	-10.54	-1.29
A-U	0.01	-0.11	0.06	0.04	-10.5	-2.34
G-C	-0.14	-0.18	0.07	0.19	-10.52	-1.3
C-G	0.14	-0.18	0.07	-0.19	-10.5	-1.27
U-A	-0.01	-0.11	0.06	-0.03	-10.51	-2.33

	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
A-U	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-4.23	-0.03	2.6	21.93	-0.19	33.72
A-U	-4.69	0.14	2.5	23.39	-0.16	31.69
G-C	-4.68	-0.14	2.5	23.26	0.21	31.83
C-G	-4.01	0	2.64	21.34	0.01	34.6
U-A	-4.68	0.14	2.5	23.27	-0.22	31.84

Table S5.4. Deviation from target and RMSD to target of helix parameters of a representative base pair in the final accepted set of structures (Table 5.1 in the main text, GP-set, 0.05 ppm, N=956)

Helix parameter of base pair 3	Deviation from target and RMSD to target[$]	Deviation from target and RMSD to target as z-score[$]
Shear (S_x) [#]	-0.01 (0.00) Å	-0.02 (0.00)
Stretch (S_y) [#]	-0.09 (0.00) Å	-0.50 (0.00)
Stagger (S_z)	0.04 (0.40) Å	0.11 (1.14)
Buckle (κ)	-1.23 (1.95) °	-0.18 (0.29)
Prop. Twist (ω)	2.71 (3.77) °	0.38 (0.53)
Opening (σ)[#]	3.08 (0.00) °	0.72 (0.00)
X-displacement (D_x)	0.09 (0.50) Å	0.08 (0.42)
Y-displacement (D_Y)	-0.06 (0.53) Å	-0.06 (0.56)
h-Rise (D_z)	0.11 (0.23) Å	0.27 (0.56)
Inclination (η)	-2.93 (5.01) °	-0.35 (0.59)
Tip (θ)	0.73 (4.41) °	0.10 (0.61)
h-Twist (Ω)	-1.09 (2.58) °	-0.31 (0.72)

[$] Deviation to target: for each helix parameter the average value over the set of converged structures was calculated and the corresponding value in the target helix subtracted. In parentheses, the standard deviation of each of the helix parameters with respect to target helix value is given. In column 2, the values are given in Å and°, while in column 3 they have been normalized against σ_{helixpar} as given in Table S5.1. [#] The 3 helix parameters marked with # were kept fixed at their standard helix values (Table S5.1) during structure calculations, because they are insensitive to chemical shift changes.

Table S5.5. Mean deviation from *target structure* and standard deviations from *target structure* of all helix parameters of the 956 accepted structures, obtained using a chemical shift tolerance of 0.05 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.01	0.00	-0.09	0.00	0.09	0.28	3.10	3.20	0.79	5.21	3.10	0.00
C-G	-0.14	0.00	-0.02	0.00	-0.07	0.40	-1.90	2.39	-0.61	4.14	2.03	0.00
A-U	-0.01	0.00	-0.09	0.00	0.04	0.40	-1.23	1.95	2.71	3.77	3.08	0.00
G-C	0.14	0.00	-0.02	0.00	-0.04	0.42	0.28	1.88	4.22	5.08	2.04	0.00
C-G	-0.14	0.00	-0.02	0.00	0.02	0.41	-0.02	1.58	4.41	5.60	2.01	0.00
U-A	0.01	0.00	-0.09	0.00	-0.29	0.50	2.12	2.38	2.23	3.43	3.07	0.00
Mean	-0.03	0.00	-0.06	0.00	-0.04	0.40	0.39	2.23	2.29	4.54	2.56	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.35	0.61	0.16	0.45	0.06	0.10	1.70	2.90	-1.28	5.02	-0.14	1.59
A-U	0.09	0.50	-0.06	0.53	0.11	0.23	-2.93	5.01	0.73	4.41	-1.09	2.58
G-C	-0.04	0.58	0.00	0.37	0.00	0.17	-0.20	3.47	-0.27	3.65	-0.47	2.53
C-G	-0.73	0.58	-0.08	0.35	0.02	0.11	1.36	2.25	0.76	3.75	-3.22	2.06
U-A	-0.08	0.43	0.33	0.37	-0.06	0.11	-0.13	2.40	-3.63	3.90	-0.83	1.74
Mean	-0.22	0.54	0.07	0.41	0.03	0.14	-0.04	3.21	-0.74	4.15	-1.15	2.10

Table S5.6. Mean deviation from *target structure* and root mean square deviation from *target structure* of all helix parameters of the 956 accepted structures, obtained in the convergence check using a chemical shift tolerance of 0.05 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts; means and standard deviations are given here normalized to σ_{helixpar} the standard deviation of allowed A-helix space, given in Table S5.1.

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.02	0.00	-0.50	0.00	0.26	0.80	0.46	0.47	0.11	0.74	0.73	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.20	1.14	-0.28	0.35	-0.09	0.59	0.48	0.00
A-U	-0.02	0.00	-0.50	0.00	0.11	1.14	-0.18	0.29	0.38	0.53	0.72	0.00
G-C	0.33	0.00	-0.11	0.00	-0.11	1.20	0.04	0.28	0.60	0.72	0.48	0.00
C-G	-0.33	0.00	-0.11	0.00	0.06	1.17	0.00	0.23	0.62	0.79	0.47	0.00
U-A	0.02	0.00	-0.50	0.00	-0.83	1.43	0.31	0.35	0.32	0.49	0.72	0.00
Mean	-0.06	0.00	-0.31	0.00	-0.12	1.15	0.06	0.33	0.32	0.64	0.60	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.29	0.51	0.17	0.48	0.15	0.24	0.20	0.34	-0.18	0.70	-0.04	0.45
A-U	0.08	0.42	-0.06	0.56	0.27	0.56	-0.35	0.59	0.10	0.61	-0.31	0.72
G-C	-0.03	0.49	0.00	0.39	0.00	0.41	-0.02	0.41	-0.04	0.51	-0.13	0.71
C-G	-0.61	0.49	-0.09	0.37	0.05	0.27	0.16	0.27	0.11	0.52	-0.90	0.58
U-A	-0.07	0.36	0.35	0.39	-0.15	0.27	-0.02	0.28	-0.50	0.54	-0.23	0.49
Mean	-0.18	0.45	0.07	0.44	0.06	0.35	-0.01	0.38	-0.10	0.58	-0.32	0.59

Table S5.7. Mean deviation from *target structure* and standard deviations from *target structure* of all helix parameters of the 956 accepted structures, obtained using a chemical shift tolerance of 0.10 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.01	0.00	-0.09	0.00	0.01	0.39	3.21	3.98	1.35	6.64	3.10	0.00
C-G	-0.14	0.00	-0.02	0.00	-0.09	0.48	-2.40	3.36	0.07	5.71	2.03	0.00
A-U	-0.01	0.00	-0.09	0.00	0.01	0.48	-1.13	2.71	3.07	5.20	3.08	0.00
G-C	0.14	0.00	-0.02	0.00	-0.07	0.51	0.23	2.77	4.25	6.83	2.04	0.00
C-G	-0.14	0.00	-0.02	0.00	0.00	0.50	-0.15	2.51	4.68	7.61	2.01	0.00
U-A	0.01	0.00	-0.09	0.00	-0.31	0.57	1.96	3.53	2.76	4.93	3.07	0.00
Mean	-0.03	0.00	-0.06	0.00	-0.08	0.49	0.29	3.14	2.70	6.15	2.56	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.45	0.79	0.1	0.6	0.08	0.16	1.86	3.79	-0.61	6.5	-0.33	2.32
A-U	0.08	0.68	-0.02	0.73	0.13	0.31	-3.67	6.87	0.4	6	-1.62	3.45
G-C	-0.11	0.81	-0.02	0.52	0.01	0.24	-0.32	4.53	-0.32	5.23	-0.61	3.51
C-G	-0.77	0.83	-0.1	0.52	0.03	0.18	1.09	3.64	1.03	5.38	-3.37	2.78
U-A	-0.14	0.6	0.38	0.54	-0.05	0.18	-0.08	3.43	-4.02	5.54	-0.97	2.46
Mean	-0.28	0.74	0.07	0.58	0.04	0.21	-0.22	4.45	-0.70	5.73	-1.38	2.90

Table S5.8. Mean deviation from *target structure* and root mean square deviation from *target structure* of all helix parameters of the 956 accepted structures, obtained in the convergence check using a chemical shift tolerance of 0.10 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts; means and standard deviations are given here normalized to $\sigma_{helixpar}$ the standard deviation of allowed A-helix space, given in Table S5.1

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.02	0.00	-0.50	0.00	0.03	1.11	0.47	0.58	0.19	0.94	0.73	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.26	1.37	-0.35	0.49	0.01	0.81	0.48	0.00
A-U	-0.02	0.00	-0.50	0.00	0.03	1.37	-0.17	0.40	0.43	0.74	0.72	0.00
G-C	0.33	0.00	-0.11	0.00	-0.20	1.46	0.03	0.41	0.60	0.97	0.48	0.00
C-G	-0.33	0.00	-0.11	0.00	0.00	1.43	-0.02	0.37	0.66	1.08	0.47	0.00
U-A	0.02	0.00	-0.50	0.00	-0.89	1.63	0.29	0.52	0.39	0.70	0.72	0.00
Mean	-0.06	0.00	-0.31	0.00	-0.22	1.40	0.04	0.46	0.38	0.87	0.60	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.38	0.66	0.11	0.64	0.20	0.39	0.22	0.45	-0.08	0.90	-0.09	0.65
A-U	0.07	0.57	-0.02	0.78	0.32	0.76	-0.43	0.81	0.06	0.83	-0.46	0.97
G-C	-0.09	0.68	-0.02	0.55	0.02	0.59	-0.04	0.54	-0.04	0.73	-0.17	0.99
C-G	-0.65	0.70	-0.11	0.55	0.07	0.44	0.13	0.43	0.14	0.75	-0.95	0.78
U-A	-0.12	0.50	0.40	0.57	-0.12	0.44	-0.01	0.41	-0.56	0.77	-0.27	0.69
Mean	-0.23	0.62	0.07	0.62	0.10	0.52	-0.03	0.53	-0.10	0.80	-0.39	0.82

Table S5.9. Mean deviation from *target structure* and standard deviations from *target structure* of all helix parameters of the 955 accepted structures, obtained using a chemical shift tolerance of 0.15 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.01	0.00	-0.09	0.00	-0.04	0.46	3.32	4.49	1.68	7.73	3.10	0.00
C-G	-0.14	0.00	-0.02	0.00	-0.09	0.52	-2.68	4.05	0.54	6.71	2.03	0.00
A-U	-0.01	0.00	-0.09	0.00	-0.01	0.52	-1.07	3.35	3.22	6.21	3.08	0.00
G-C	0.14	0.00	-0.02	0.00	-0.09	0.56	0.13	3.63	4.23	8.00	2.04	0.00
C-G	-0.14	0.00	-0.02	0.00	-0.02	0.56	-0.26	3.36	4.70	8.70	2.01	0.00
U-A	0.01	0.00	-0.09	0.00	-0.32	0.61	1.88	4.51	3.29	6.02	3.07	0.00
Mean	-0.03	0.00	-0.06	0.00	-0.10	0.54	0.22	3.90	2.94	7.23	2.56	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.51	0.94	0.06	0.71	0.09	0.2	1.87	4.56	-0.11	7.57	-0.44	2.82
A-U	0.08	0.81	0.01	0.88	0.13	0.38	-4.06	7.98	0.04	7.2	-1.82	3.89
G-C	-0.14	0.95	-0.04	0.65	0.03	0.3	-0.58	5.69	-0.27	6.28	-0.75	4.09
C-G	-0.76	1	-0.1	0.67	0.04	0.26	0.83	5.01	0.97	6.67	-3.41	3.26
U-A	-0.21	0.76	0.41	0.67	-0.06	0.25	-0.01	4.5	-4.15	6.68	-1.13	2.97
Mean	-0.31	0.89	0.07	0.72	0.05	0.28	-0.39	5.55	-0.70	6.88	-1.51	3.41

Table S5.10. Mean deviation from *target structure* and root mean square deviation from *target structure* of all helix parameters of the 955 accepted structures, obtained in the convergence check using a chemical shift tolerance of 0.15 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts; means and standard deviations are given here normalized to $\sigma_{\mathrm{helixpar}}$ the standard deviation of allowed A-helix space, given in Table S5.1

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.02	0.00	-0.50	0.00	-0.11	1.31	0.49	0.66	0.24	1.09	0.73	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.26	1.49	-0.39	0.59	0.08	0.95	0.48	0.00
A-U	-0.02	0.00	-0.50	0.00	-0.03	1.49	-0.16	0.49	0.46	0.88	0.72	0.00
G-C	0.33	0.00	-0.11	0.00	-0.26	1.60	0.02	0.53	0.60	1.13	0.48	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.06	1.60	-0.04	0.49	0.66	1.23	0.47	0.00
U-A	0.02	0.00	-0.50	0.00	-0.91	1.74	0.28	0.66	0.47	0.85	0.72	0.00
Mean	-0.06	0.00	-0.31	0.00	-0.27	1.54	0.03	0.57	0.42	1.02	0.60	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.43	0.79	0.06	0.76	0.22	0.49	0.22	0.54	-0.02	1.05	-0.12	0.79
A-U	0.07	0.68	0.01	0.94	0.32	0.93	-0.48	0.94	0.01	1.00	-0.51	1.09
G-C	-0.12	0.80	-0.04	0.69	0.07	0.73	-0.07	0.67	-0.04	0.87	-0.21	1.15
C-G	-0.64	0.84	-0.11	0.71	0.10	0.63	0.10	0.59	0.13	0.93	-0.96	0.92
U-A	-0.18	0.64	0.44	0.71	-0.15	0.61	0.00	0.53	-0.58	0.93	-0.32	0.83
Mean	-0.26	0.75	0.07	0.76	0.11	0.68	-0.05	0.65	-0.10	0.96	-0.42	0.96

Table S5.11. Mean deviation from *target structure* and standard deviations from *target structure* all helix parameters of the 963 accepted structures, obtained using a chemical shift tolerance of 0.30 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.01	0.00	-0.09	0.00	-0.11	0.59	3.45	6.11	1.86	10.19	3.10	0.01
C-G	-0.14	0.00	-0.02	0.00	-0.10	0.61	-3.01	5.92	1.15	9.23	2.03	0.01
A-U	-0.01	0.00	-0.09	0.00	-0.05	0.60	-1.17	6.17	3.30	8.62	3.08	0.01
G-C	0.14	0.00	-0.02	0.00	-0.10	0.65	-0.10	6.43	4.00	10.74	2.04	0.01
C-G	-0.14	0.00	-0.02	0.00	-0.04	0.66	-0.14	6.02	4.83	10.92	2.01	0.01
U-A	0.01	0.00	-0.09	0.00	-0.30	0.70	1.24	7.10	4.54	9.17	3.07	0.01
Mean	-0.03	0.00	-0.06	0.00	-0.12	0.64	0.05	6.29	3.28	9.81	2.56	0.01

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.54	1.30	0.00	0.99	0.10	0.37	1.71	6.77	0.71	9.90	-0.50	4.18
A-U	0.13	1.20	0.03	1.17	0.17	0.53	-5.38	10.69	-0.43	9.62	-2.04	4.89
G-C	-0.13	1.30	-0.04	0.99	0.05	0.48	-1.27	8.80	-0.13	9.12	-1.00	5.25
C-G	-0.70	1.35	-0.08	1.03	0.02	0.47	0.35	8.08	0.78	9.74	-3.37	4.59
U-A	-0.29	1.17	0.35	1.03	-0.01	0.47	-0.35	8.12	-3.45	9.71	-1.31	4.26
Mean	-0.31	1.26	0.05	1.04	0.07	0.46	-0.99	8.49	-0.50	9.62	-1.64	4.63

Table S5.12. Mean deviation from *target structure* and root mean square deviation from *target structure* of all helix parameters of the 963 accepted structures, obtained in the convergence check using a chemical shift tolerance of 0.30 ppm and using the Giessner-Prettre parameterization to calculate the proton chemical shifts; means and standard deviations are given here normalized to $\sigma_{\mathrm{helixpar}}$ the standard deviation of allowed A-helix space, given in Table S5.1

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.02	0.00	-0.50	0.00	-0.31	1.69	0.51	0.90	0.26	1.44	0.73	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.29	1.74	-0.44	0.87	0.16	1.31	0.48	0.00
A-U	-0.02	0.00	-0.50	0.00	-0.14	1.71	-0.17	0.91	0.47	1.22	0.72	0.00
G-C	0.33	0.00	-0.11	0.00	-0.29	1.86	-0.01	0.94	0.57	1.52	0.48	0.00
C-G	-0.33	0.00	-0.11	0.00	-0.11	1.89	-0.02	0.88	0.68	1.54	0.47	0.00
U-A	0.02	0.00	-0.50	0.00	-0.86	2.00	0.18	1.04	0.64	1.30	0.72	0.00
Mean	-0.06	0.00	-0.31	0.00	-0.33	1.81	0.01	0.92	0.46	1.39	0.60	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	-0.45	1.09	0.00	1.05	0.24	0.90	0.20	0.80	0.10	1.38	-0.14	1.17
A-U	0.11	1.01	0.03	1.24	0.41	1.29	-0.64	1.26	-0.06	1.34	-0.57	1.37
G-C	-0.11	1.09	-0.04	1.05	0.12	1.17	-0.15	1.04	-0.02	1.27	-0.28	1.47
C-G	-0.59	1.13	-0.09	1.10	0.05	1.15	0.04	0.96	0.11	1.35	-0.95	1.29
U-A	-0.24	0.98	0.37	1.10	-0.02	1.15	-0.04	0.96	-0.48	1.35	-0.37	1.20
Mean	-0.26	1.06	0.05	1.11	0.16	1.13	-0.12	1.00	-0.07	1.34	-0.46	1.30

Table S5.13. Mean deviation from *target structure* and standard deviations from *target structure* all helix parameters of the 900 accepted structures, obtained using a chemical shift tolerance of 0.10 ppm and using the Case (DC) ring current parameter to calculate the proton chemical shifts

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.01	0.00	-0.09	0.00	0.39	0.62	0.59	5.25	-5.15	9.79	3.10	0.00
C-G	-0.14	0.00	-0.02	0.00	0.47	0.44	-2.81	5.76	3.01	7.59	2.03	0.00
A-U	-0.01	0.00	-0.09	0.00	-0.22	0.53	-5.59	7.87	-3.90	5.39	3.08	0.00
G-C	0.14	0.00	-0.02	0.00	-0.03	0.59	-7.03	7.49	-10.13	9.10	2.04	0.00
C-G	-0.14	0.00	-0.02	0.00	0.24	0.59	-0.13	6.76	-6.18	9.08	2.01	0.00
U-A	0.01	0.00	-0.09	0.00	-0.42	0.55	-4.34	7.72	0.94	7.36	3.07	0.00
Mean	-0.03	0.00	-0.06	0.00	0.07	0.55	-3.22	6.81	-3.57	8.05	2.56	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	0.08	0.93	-0.42	0.60	0.03	0.41	1.98	8.88	4.25	6.83	0.42	2.90
A-U	-0.60	0.77	0.49	0.77	-0.42	0.54	1.49	9.72	-7.18	7.41	-0.58	3.92
G-C	0.40	0.66	0.22	0.65	-0.17	0.37	3.20	4.99	-2.12	8.03	0.88	3.85
C-G	0.50	0.75	-0.07	0.50	-0.54	0.34	3.69	4.06	1.39	7.03	1.04	3.11
U-A	-0.07	0.69	0.18	0.52	-0.30	0.40	3.39	5.82	-6.66	6.39	4.46	2.55
Mean	0.06	0.76	0.08	0.61	-0.28	0.41	2.75	6.69	-2.06	7.14	1.24	3.27

Table S5.14. Mean deviation from *target structure* and root mean square deviation from *target structure* of all helix parameters of the 900 accepted structures, obtained in the convergence check using a chemical shift tolerance of 0.10 ppm and using the Case (DC) ring current parameter set to calculate the proton chemical shifts; means and standard deviations are given here normalized to σ_{helixpar} the standard deviation of allowed A-helix space, given in Table S5.1

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	-0.02	0.00	-0.50	0.00	1.11	1.77	0.09	0.77	-0.73	1.38	0.73	0.00
C-G	-0.33	0.00	-0.11	0.00	1.34	1.26	-0.41	0.85	0.43	1.07	0.48	0.00
A-U	-0.02	0.00	-0.50	0.00	-0.63	1.51	-0.82	1.16	-0.55	0.76	0.72	0.00
G-C	0.33	0.00	-0.11	0.00	-0.09	1.69	-1.03	1.10	-1.43	1.29	0.48	0.00
C-G	-0.33	0.00	-0.11	0.00	0.69	1.69	-0.02	0.99	-0.87	1.28	0.47	0.00
U-A	0.02	0.00	-0.50	0.00	-1.20	1.57	-0.64	1.13	0.13	1.04	0.72	0.00
Mean	-0.06	0.00	-0.31	0.00	0.20	1.58	-0.47	1.00	-0.50	1.14	0.60	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
A-U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C-G	0.07	0.78	-0.45	0.64	0.07	1.00	0.23	1.05	0.59	0.95	0.12	0.81
A-U	-0.50	0.65	0.52	0.82	-1.02	1.32	0.18	1.15	-1.00	1.03	-0.16	1.10
G-C	0.34	0.55	0.23	0.69	-0.41	0.90	0.38	0.59	-0.29	1.12	0.25	1.08
C-G	0.42	0.63	-0.07	0.53	-1.32	0.83	0.44	0.48	0.19	0.98	0.29	0.87
U-A	-0.06	0.58	0.19	0.55	-0.73	0.98	0.40	0.69	-0.93	0.89	1.25	0.72
Mean	0.05	0.64	0.08	0.65	-0.68	1.01	0.33	0.79	-0.29	0.99	0.35	0.92

Table S5.15. Structural characteristics of the chemical shift derived accepted helix structures as function of the number of chemical shift restraints per basepair employed during structure derivation

Chem. Shifts per basepair	$\mathrm{RMSD}_{\mathrm{av2target}}$ (Å)	$\mathrm{RMSD}_{\mathrm{pairwise}}$ (Å)	RMSD (chem. shift) (ppm)	VCSQ	N
15.4	0.24	0.5 ± 0.2	0.089 ± 0.004	0.137 ± 0.006	456
15.0	0.56	0.5 ± 0.2	0.067 ± 0.003	0.144 ± 0.006	467
14.0	0.53	0.5 ± 0.2	0.082 ± 0.015	0.179 ± 0.032	291
13.0	0.41	0.5 ± 0.2	0.101 ± 0.022	0.223 ± 0.047	227
12.0	0.43	0.6 ± 0.2	0.116 ± 0.028	0.256 ± 0.059	130
11.0	0.43	0.8 ± 0.2	0.131 ± 0.022	0.286 ± 0.048	54
10.0	0.55	0.9 ± 0.3	0.141 ± 0.026	0.308 ± 0.056	36
9.0	0.48	1.0 ± 0.3	0.152 ± 0.017	0.329 ± 0.036	30
8.0	0.52	1.0 ± 0.3	0.157 ± 0.024	0.342 ± 0.051	32
7.4[#]	0.85[#]	0.9 ± 0.3[#]	0.084 ± 0.010[#]	0.183 ± 0.021[#]	230[#]
7.4[$]	0.52[$]	1.0 ± 0.3[$]	0.081 ± 0.007[$]	0.175 ± 0.016[$]	175[$]
7.0	0.70	1.1 ± 0.3	0.177 ± 0.030	0.387 ± 0.065	16
6.0	0.72	1.2 ± 0.3	0.188 ± 0.025	0.408 ± 0.055	101
5.4[&]	0.83[&]	1.0 ± 0.3[&]	0.106 ± 0.018[&]	0.232 ± 0.038[&]	113[&]
5.0	0.89	1.4 ± 0.4	0.210 ± 0.024	0.454 ± 0.052	21
4.0	0.98	1.5 ± 0.3	0.233 ± 0.033	0.506 ± 0.071	12
3.0	1.26	1.6 ± 0.3	0.254 ± 0.032	0.551 ± 0.069	9

The structural characteristics of the accepted structures are shown, i.e. converged SVD structures that are backbone regularized and finally checked violations: no chemical shift violations >0.1 ppm, no H-bond distance violations, no distance violations > 0.5 Å from standard helix NOE values. In the columns from left to right, the following data are displayed: The number of chemical shifts used per basepair to determine a 10-basepair helix structure (the first row includes all chemical shifts). The RMSD of the average accepted structure to the target structure (after backbone regularization & violation check). The pairwise RMSD of the central six basepairs of the accepted structures. The RMSD of the chemical shifts to the target values for all chemical shifts of the accepted structures. The VCSQ-factor (see main text) of the accepted structures. The number of structures that converged and were accepted after backbone regularization, i.e. no chemical shift violations >0.1 ppm, no H-bond distance violations, no distance violations > 0.5 Å from standard helix NOE values. Special sets of easy measurable chemical shift selections are indicated with symbols: [#] H1’, H2’ and aromatic proton chemical shifts [$] H1’, H4’, and aromatic proton chemical shifts [&] H1’ and aromatic proton chemical shifts

Table S5.16a. Helix parameters of the deposited NMR stem-loop structure HIV1_RNA (entry: 1PJY)

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.42	-0.20	0.18	0.02	0.35	-1.01	6.81	-12.59	7.07	0.74	4.27
G-C	-0.41	0.36	-0.27	0.13	-0.35	0.45	-8.04	8.99	0.12	7.66	-3.09	2.50
G-C	-0.81	0.24	-0.38	0.15	-0.20	0.38	0.17	6.46	0.50	8.92	0.22	1.61
C-G	0.55	0.36	-0.31	0.14	-0.26	0.38	1.15	6.07	5.15	6.86	-1.12	1.68
C-G	0.43	0.42	-0.32	0.14	-0.54	0.32	1.48	5.84	0.36	7.73	0.02	1.80
U-A	0.17	0.28	-0.30	0.12	-0.30	0.46	-6.98	2.07	3.19	4.96	0.38	3.55
U-A	0.14	0.30	-0.26	0.05	-0.28	0.32	0.17	4.28	1.94	4.46	-0.55	2.94
C-G	0.03	0.46	-0.21	0.10	0.25	0.37	4.98	4.24	7.68	6.24	2.21	1.92
C-G	0.65	0.39	-0.34	0.11	0.04	0.34	3.55	5.17	0.61	8.44	-0.17	2.52
C-G	0.90	0.28	-0.47	0.10	-0.08	0.36	-2.70	5.20	-6.30	6.85	-1.26	1.67
A-A	6.40	0.20	-4.05	0.39	0.62	0.44	6.34	5.92	6.02	10.71	-3.26	12.05

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	-4.30	1.19	-0.10	0.94	2.71	0.41	16.55	8.46	0.07	7.19	33.23	3.56
G-C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-3.54	0.82	-1.40	0.70	2.69	0.49	19.50	7.94	7.38	5.32	33.14	2.43
C-G	-2.59	0.94	0.15	0.67	3.53	0.58	8.58	7.68	1.31	4.68	36.73	2.95
C-G	-3.80	1.32	0.38	0.66	3.53	0.58	6.99	7.64	-4.96	4.53	29.58	3.34
U-A	-4.97	0.90	-0.22	0.91	3.50	0.67	13.49	7.70	2.17	4.29	28.99	2.86
U-A	-4.18	1.22	0.38	0.80	2.69	0.27	17.91	6.16	-2.44	1.75	31.36	2.27
C-G	-3.82	0.95	-0.53	0.63	3.20	0.40	12.01	5.83	4.93	3.96	30.15	2.86
C-G	-3.77	0.85	0.72	0.74	2.86	0.52	22.29	9.87	-5.27	3.73	36.90	5.87
C-G	-3.58	0.57	0.06	0.59	2.90	0.47	26.52	6.11	-5.88	3.18	40.66	2.74
A-A	-0.82	0.54	0.50	0.85	3.04	0.13	12.06	5.14	0.42	3.15	52.32	2.77

Deviations in the helix parameters from standard A-helix values of 2-2.3$\sigma_{helixpar}$, 2.3-3.0$\sigma_{helixpar}$, and more than 3.0$\sigma_{helixpar}$ are marked as bold, bolded and underlined, and italic respectively. The standard A-helix values and their STDEVs ($\sigma_{helixpar}$) are based on X-ray structures (Table S5.1) and given in the top row. The $\sigma_{helixpar}$ values indicate allowed A-helix conformational space. Marked in grey are basepairs that show end-effects due to proximity to capping hairpin loops.

Table S5.16b. Helix Parameters of the deposited NMR stem-loop structure of SIV_RNA (entry: 2JTP)

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.42	-0.20	0.18	0.02	0.35	-1.01	6.81	-12.59	7.07	0.74	4.27
G-C	0.00	0.37	-0.11	0.06	-0.16	0.10	3.72	0.83	3.50	2.91	-1.16	0.94
G-C	-0.91	0.03	-0.34	0.01	0.01	0.10	2.40	1.09	0.13	2.23	-0.07	0.72
A-U	-0.03	0.13	-0.07	0.13	-0.26	0.11	-4.90	1.45	-6.47	3.98	-8.04	1.97
U-A	-0.85	0.20	-0.12	0.04	-0.18	0.11	-1.56	1.63	-7.02	2.47	-8.39	2.20
G-C	0.25	0.15	-0.11	0.03	-0.43	0.15	-0.54	2.09	1.31	3.44	-1.12	1.34
G-C	-0.35	0.16	-0.16	0.06	0.08	0.13	-3.51	2.40	-4.29	3.04	-2.79	0.82
G-C	-0.68	0.17	-0.25	0.07	-0.29	0.12	-5.10	2.64	-6.48	2.95	-0.74	1.12
G-C	0.34	0.06	-0.15	0.03	-0.52	0.20	-5.33	2.15	-10.59	2.77	2.48	1.75
A-U	0.53	0.29	-0.02	0.05	0.26	0.13	-7.12	1.84	-3.11	4.31	1.45	3.50
A-U	0.32	0.32	-0.16	0.07	-0.30	0.12	-0.16	2.06	-6.04	4.78	3.25	3.48
A-U	0.32	0.20	-0.10	0.04	0.08	0.19	-3.06	3.15	2.05	5.49	5.72	2.93
G-A	6.28	1.34	-3.38	0.99	-0.45	0.54	-4.77	5.25	-1.53	10.22	3.93	11.20

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	-4.30	1.19	-0.10	0.94	2.71	0.41	16.55	8.46	0.07	7.19	33.23	3.56
G-C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-4.88	0.70	-0.89	0.41	3.12	0.32	21.12	4.29	5.80	1.40	32.49	3.03
A-U	-5.26	0.62	1.92	0.30	2.11	0.63	31.82	6.34	-2.59	1.18	36.42	1.12
U-A	-0.97	2.07	0.71	0.28	3.84	0.47	-10.50	11.21	4.07	1.51	29.19	2.49
G-C	-3.16	0.69	-0.80	0.43	3.20	0.50	23.17	6.70	0.49	1.74	37.80	2.23
G-C	-2.95	0.92	0.75	0.36	3.83	0.35	16.75	5.96	6.70	1.75	32.26	1.59
G-C	-4.16	0.47	-1.51	0.25	2.25	0.27	32.02	4.26	-0.22	2.14	37.25	2.08
G-C	-4.29	0.78	0.32	0.29	2.11	0.37	32.29	7.18	-2.09	1.74	37.58	2.45
A-U	-1.06	1.25	-0.29	0.96	3.70	0.58	5.30	9.13	9.69	3.66	38.66	2.31
A-U	-4.72	1.59	-0.60	0.84	2.06	0.81	23.15	12.45	-3.01	1.30	29.73	2.47
A-U	0.76	2.07	-0.43	0.38	4.10	0.63	-15.58	20.34	5.80	3.36	39.80	3.16
G-A	-1.93	0.88	0.67	1.38	2.93	0.44	13.75	6.84	-5.44	4.97	57.50	6.04

Deviations in the helix parameters from standard A-helix values of 2-2.3$\sigma_{helixpar}$, 2.3-3.0$\sigma_{helixpar}$, and more than 3.0$\sigma_{helixpar}$ are marked as bold, bold and underlined, and italic, respectively. The standard A-helix values and their STDEVs ($\sigma_{helixpar}$) are based on X-ray structures (Table S5.1) and given in the top row. The $\sigma_{helixpar}$ values indicate allowed A-helix conformational space. Marked in grey are basepairs that show end-effects due to proximity to capping hairpin loops.

Table S5.17a. Statistics of the chemical shift optimized structure of HIV1_RNA (10 basepairs), based on 87 experimental chemical shifts from BMRB.

1PJY	1PJY, deposited NMR structure (20 struct.)	Start (500 struct.)	SVD (336 struct.)
$\mathrm{RMSD}_{\mathrm{chemshift}}$(all) [ppm]	0.16 ± 0.01	4 ± 56	0.056 ± 0.003
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (base) [ppm]	0.20 ± 0.02	1.4 ± 6.9	0.040 ± 0.005
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (sugar) [ppm]	0.14 ± 0.01	4.3 ± 68.6	0.062 ± 0.003
$\mathrm{RMSD}_{\mathrm{pairwise}}$ (all) [Å]	1.5 ± 0.5	3.3 ± 0.7	0.9 ± 0.3
$\mathrm{RMSD}_{\mathrm{pairwise}}$(base) [Å]	1.4 ± 0.5	2.8 ± 0.6	0.8 ± 0.3
$\mathrm{RMSD}_{\mathrm{pairwise}}$(sugar) [Å]	1.6 ± 0.5	3.6 ± 0.7	1.0 ± 0.4
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$(all) [Å]	-	2.4	1.5
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$(base) [Å]	-	2.4	1.3
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$(sugar) [Å]	-	2.3	1.7

RMSDavgto1PJY is the RMSD between the coordinate averaged final accepted structure and the average structure available in the PDB (entry code: 1PJY).

Table S5.17b. Statistics of the chemical shift optimized structure of SIV_RNA (12 basepairs), based on 108 experimental chemical shifts from BMRB.

2JTP	2JTP, deposited NMR structure (20 struct.)	Start (500 struct.)	SVD (189 struct.)
$RMSD_{chemshift}$ (all) [ppm]	0.27 ± 0.02	1.6 ± 3.6	0.051 ± 0.004
$RMSD_{chemshift}$ (base) [ppm]	0.31 ± 0.03	1.2 ± 1.7	0.039 ± 0.007
$RMSD_{chemshift}$ (sugar) [ppm]	0.24 ± 0.01	1.5 ± 4.4	0.056 ± 0.004
$RMSD_{RDC}$ [Hz]	9.8 ± 1.8	27.0± 12.5	
$RMSD_{pairwise}$ (all) [Å]	0.9 ± 0.3	3.6 ± 0.7	1.4 ± 0.4
$RMSD_{pairwise}$(base) [Å]	0.9 ± 0.3	3.2 ± 0.7	1.2 ± 0.3
$RMSD_{pairwise}$(sugar) [Å]	0.9 ± 0.3	4.0 ± 0.8	1.6 ± 0.4
$RMSD_{avgto2JTP}$ (all) [Å]	-	2.2	1.7
$RMSD_{avgto2JTP}$ (base) [Å]	-	2.3	1.9
$RMSD_{avgto2JTP}$ (sugar) [Å]	-	2.1	1.3

$RMSD_{avgto2JTP}$ is the RMSD between the coordinate averaged final accepted structure and the average structure available in the PDB (entry code: 2JTP).

Table S5.18a. Statistics of the final accepted structures of the HIV1_RNA (10 basepairs), based on 87 experimental chemical shifts from BMRB. The number of distance restraints in the standard NOE set is 176.

1PJY	Regularized (65 struct.)
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (all) [ppm]	0.118 ± 0.003
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (base) [ppm]	0.062 ± 0.008
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (sugar) [ppm]	0.137 ± 0.004
$\mathrm{RMSD}_{\mathrm{NOE}}$ (standard) [Å]	0.084 ± 0.005
$\mathrm{RMSD}_{\mathrm{pairwise}}$ (all) [Å]	0.7 ± 0.3
$\mathrm{RMSD}_{\mathrm{pairwise}}$(base) [Å]	0.6 ± 0.2
$\mathrm{RMSD}_{\mathrm{pairwise}}$(sugar) [Å]	0.8 ± 0.3
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$ (all) [Å]	1.4
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$ (base) [Å]	1.6
$\mathrm{RMSD}_{\mathrm{avgto1PJY}}$ (sugar) [Å]	1.3

Table S5.18b. Statistics of the final accepted structures of the SIV_RNA (12 basepairs), based on 108 experimental chemical shifts from BMRB. The number of distance restraints in the standard NOE set is 209, and in the published NOE set 403. The number of RDC restraints is 24; 4 sugar, and 20 aromatic bond vectors.

2JTP	Reg. (37 struct.) NOE-set: pub. RDCs excluded	Reg. (16 struct.) NOE-set: pub. RDCs included	Reg. (25 struct. NOE-set: standard RDCs excluded	Reg. (10 struct.) NOE-set: standard RDCs included
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (all) [ppm]	0.080 ± 0.008	0.079 ± 0.009	0.079 ± 0.006	0.088 ± 0.005
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (base) [ppm]	0.047 ± 0.013	0.053 ± 0.017	0.059 ± 0.015	0.081 ± 0.009
$\mathrm{RMSD}_{\mathrm{chemshift}}$ (sugar) [ppm]	0.092 ± 0.008	0.089 ± 0.009	0.087 ± 0.006	0.091 ± 0.006
$\mathrm{RMSD}_{\mathrm{NOE}}$ [Å]	0.1 ± 0.017	0.1 ± 0.014	0.107 ± 0.034	0.102 ± 0.040
$\mathrm{RMSD}_{\mathrm{RDC}}$ [Hz]	18.9 ± 3.8	12.6 ± 3.3	23.0 ± 3.2	14.0 ± 3.0
$\mathrm{RMSD}_{\mathrm{pairwise}}$ (all) [Å]	1.3 ± 0.4	1.2 ± 0.3	1.2 ± 0.3	1.2 ± 0.3
$\mathrm{RMSD}_{\mathrm{pairwise}}$(base) [Å]	1.1 ± 0.3	1.0 ± 0.2	1.0 ± 0.3	1.0 ± 0.2
$\mathrm{RMSD}_{\mathrm{pairwise}}$(sugar) [Å]	1.5 ± 0.4	1.3 ± 0.4	1.4 ± 0.4	1.4 ± 0.4
$\mathrm{RMSD}_{\mathrm{avgto2JTP}}$ (all) [Å]	1.6	1.7	1.7	1.6
$\mathrm{RMSD}_{\mathrm{avgto2JTP}}$ (base)[Å]	1.2	1.2	1.2	1.3
$\mathrm{RMSD}_{\mathrm{avgto2JTP}}$ (sugar) [Å]	2.0	2.2	2.0	1.7

Tabulated are different combinations of restraints that were used during the backbone regularization protocol. This illustrates that the effect of different combinations of restraints is minimal for helices.

Table S5.19a. Helix parameters of the ensemble of 65 final accepted structures of frame-shift inducing RNA element from HIV1

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.42	-0.20	0.18	0.02	0.35	-1.01	6.81	-12.59	7.07	0.74	4.27
G-C	0.00	0.00	-0.20	0.00	0.35	0.44	-7.92	5.95	-3.30	6.96	0.74	0.01
G-C	0.00	0.00	-0.20	0.00	0.16	0.39	-1.06	4.77	-18.48	5.63	0.74	0.01
C-G	0.00	0.00	-0.20	0.00	0.21	0.37	7.23	6.34	-12.09	4.90	0.74	0.01
C-G	0.00	0.00	-0.20	0.00	0.21	0.32	7.95	5.68	-7.05	6.57	0.74	0.01
U-A	0.00	0.00	-0.20	0.00	-0.03	0.43	6.47	5.11	-5.51	8.58	0.74	0.01
U-A	0.00	0.00	-0.20	0.00	0.16	0.40	7.03	6.13	-13.98	6.34	0.74	0.01
C-G	0.00	0.00	-0.20	0.00	0.12	0.46	13.21	4.49	-10.58	5.32	0.74	0.01
C-G	0.00	0.00	-0.20	0.00	-0.17	0.20	16.83	1.43	4.48	3.44	0.74	0.01
C-G	0.90	0.00	-0.47	0.00	-0.08	0.00	-2.70	0.01	-6.30	0.01	-1.26	0.01
A-A	6.40	0.00	-4.05	0.00	0.62	0.00	6.34	0.01	6.02	0.01	-3.26	0.00

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	-4.30	1.19	-0.10	0.94	2.71	0.41	16.55	8.46	0.07	7.19	33.23	3.56
G-C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-4.16	1.03	0.56	0.68	2.33	0.36	25.82	7.18	0.52	6.64	27.59	3.51
C-G	-3.44	0.48	-0.28	0.45	2.59	0.42	14.04	4.28	1.77	5.10	33.10	3.15
C-G	-5.46	0.77	-0.97	0.60	2.30	0.47	27.76	6.73	1.53	5.39	27.42	3.82
U-A	-6.14	0.96	-0.28	0.69	2.51	0.44	25.52	6.56	0.18	6.49	28.12	4.56
U-A	-4.43	0.76	-0.43	0.61	2.73	0.37	19.88	6.62	4.40	5.86	30.91	4.46
C-G	-5.06	0.61	0.09	0.66	2.43	0.34	19.48	6.06	4.00	6.52	25.87	3.63
C-G	-6.62	0.60	0.06	0.65	2.24	0.42	31.00	7.27	0.70	6.39	26.02	2.72
C-G	-3.58	0.00	0.06	0.00	2.90	0.00	26.52	0.01	-5.88	0.01	40.66	0.01
A-A	-0.82	0.00	0.50	0.00	3.04	0.00	12.06	0.01	0.42	0.01	52.32	0.00

Helix parameters deviating more than 2.2σo are marked in bold; σhelixpar is the RMSD of the helix parameter indicating allowed A-helix conformational space as derived based on X-ray structures (Table S5.1). Marked in grey are basepairs that show end-effects due to proximity to capping hairpin loops.

Table S5.19b. Helix parameters of the ensemble of 16 final accepted structures of frame-shift inducing RNA element from SIV

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.42	-0.20	0.18	0.02	0.35	-1.01	6.81	-12.59	7.07	0.74	4.27
G-C	0.00	0.00	-0.20	0.00	0.23	0.57	-8.08	5.99	0.68	6.78	0.74	0.01
G-C	0.00	0.00	-0.20	0.00	0.07	0.44	1.60	5.75	-18.71	8.51	0.74	0.01
A-U	0.00	0.00	-0.20	0.00	0.09	0.44	-6.33	4.49	1.30	5.38	0.74	0.01
U-A	0.00	0.00	-0.20	0.00	-0.18	0.26	-2.18	5.14	-9.11	7.61	0.74	0.00
G-C	0.00	0.00	-0.20	0.00	-0.15	0.52	-11.26	5.22	-5.82	7.36	0.74	0.01
G-C	0.00	0.00	-0.20	0.00	-0.08	0.42	-9.16	7.16	-5.24	7.12	0.74	0.01
G-C	0.00	0.00	-0.20	0.00	-0.01	0.53	-10.31	5.70	-2.88	9.75	0.74	0.01
G-C	0.00	0.00	-0.20	0.00	0.04	0.35	-7.75	6.28	-12.51	4.70	0.74	0.01
A-U	0.00	0.00	-0.20	0.00	0.19	0.46	-10.92	4.63	-4.67	7.57	0.74	0.01
A-U	0.00	0.00	-0.20	0.00	0.57	0.36	-3.58	5.51	-13.17	3.85	0.74	0.00
A-U	0.00	0.00	-0.20	0.00	0.33	0.34	-16.91	3.92	-11.81	7.25	0.74	0.01
G-A	1.28	0.32	1.07	0.24	1.39	1.35	-8.01	10.42	1.58	10.55	-0.18	4.37

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	-4.30	1.19	-0.10	0.94	2.71	0.41	16.55	8.46	0.07	7.19	33.23	3.56
G-C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-4.59	0.66	0.34	0.77	2.13	0.43	25.98	6.67	3.11	5.91	28.58	2.86
A-U	-4.14	0.87	-0.65	0.64	2.78	0.35	22.61	3.30	3.11	6.05	28.40	3.93
U-A	-4.66	1.03	-0.71	0.80	2.85	0.43	22.43	8.03	0.16	7.78	34.33	3.52
G-C	-5.71	1.05	-0.20	0.92	2.64	0.35	30.74	6.09	2.20	7.37	28.42	5.32
G-C	-4.69	0.68	-0.42	0.85	2.79	0.54	16.95	6.12	1.03	7.09	32.21	4.29
G-C	-4.50	0.75	-0.55	0.74	2.60	0.45	22.42	9.24	1.25	7.93	33.04	3.86
G-C	-4.31	0.58	-0.49	0.54	2.57	0.54	19.07	7.56	3.84	5.79	34.37	4.47
A-U	-4.56	0.86	-0.05	0.99	2.83	0.35	18.11	7.86	-1.25	8.32	27.40	4.53
A-U	-4.39	0.91	-0.67	0.53	2.55	0.43	21.51	6.82	5.56	6.21	33.56	5.38
A-U	-6.54	0.68	0.51	0.63	2.05	0.42	38.48	5.03	-5.15	7.11	27.46	2.79
G-A	-5.34	1.00	-0.11	1.13	2.66	0.92	21.28	13.19	7.65	9.86	36.85	5.13

Structures above result from a backbone regularization protocol in which published NOEs and RDCs were included as restraints. Helix parameters deviating more than 2.0σhelixpar from A-helix values are marked in bold; σhelixpar is the RMSD of the helix parameter indicating allowed A-helix conformational space as derived based on X-ray structures (Table S5.1). Marked in grey is the basepair that show end-effects due to proximity to capping hairpin loops.

References

1. Babcock MS, Pednault EPD, Olson WK. 1994. Nucleic Acid Structure Analysis: Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable Between Structures. *J Mol Biol* 237:125-56
2. Babcock MS, Olson WK. 1994. The Effect of Mathematics and Coordinate System on Comparability and "Dependencies" of Nucleic Acid Structure Parameters. *J Mol Biol* 237:98-124
3. El Hassan MA, Calladine CR. 1995. The Assessment of the Geometry of Dinucleotide Steps in Double-Helical DNA; a New Local Calculation Scheme. *J Mol Biol* 251:648-64
4. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-21
5. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, et al. 1989. Definitions and Nomenclature of Nucleic Acid Structure Parameters. *J Mol Biol* 205:787-91
6. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37
7. Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* 10:337-50
8. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2007. *Numerical Recipes, The art of scientific computing*. Cambridge: Cambridge University Press. 1256 pp.
9. Lu X. 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* 3:1213-27
10. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. 2001. Prediction of Proton Chemical Shifts in RNA. Their use in Structure refinement and validation. *J Biomol NMR* 21:11-29
11. Ribas-Prado R, Giessner-Prettre C. 1981. Parameters for the calculation of the ring current and atomic magnetic anisotropy contributions to magnetic shielding constants: Nucleic acid bases and intercalating agents. *J.Mol.Struct. THEOCHEM* 76:81-92
12. Giessner-Prettre C, Pullman B. 1987. Quantum mechanical calculations of NMR chemical shifts in nucleic acids. *Quart. Rev. Biophys.* 20:113-72
13. Case DA. 1995. Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* 6:341-6
14. Brunger AT. 1992. *X-PLOR version 3.1 A system for X-ray Crystallography and NMR*. New Haven: Yale University Press
15. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387
16. Gelbin A, Schneider B, Clowney L, Hsieh SH, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Am Chem Soc* 118:519-29
17. Feng Z, Westbrook J, Berman HM. 1998. NUCheck. *Rep. NDB-407*, Rutgers University, New Brunswick

Chapter 6

Dickerson Dodecamer Structure Derived from Experimental Proton Chemical Shifts

Werf, R.M. van der; Tessari, M.; Wijmenga, S. S. (2009): *Manuscript in Preparation*

6 Dickerson Dodecamer Structure Derived from Experimental Proton Chemical Shifts

6.1 Abstract

Chemical shifts are highly sensitive probes of molecular structure and form the most readily and accurately measurable NMR parameters. We show here via the example of the DNA Dickerson dodecamer that DNA helix structures can be derived with high accuracy based on proton chemical shifts alone. Non-bonded force-field terms do not affect the derived structures. In addition, no use is made of sequence homology. The chemical shift derived structures show no violations with respect to experimental NOE, dihedral angle, and RDC restraints. The structures closely resemble the two deposited high-resolution X-ray structures and the deposited high-quality NMR structure, although resemblance to the X-ray structures is greater. The helix parameters of the chemical shift derived structures resemble the values in the X-ray and NMR structures, attesting to the accuracy of the structure. Even sequence specific variations seen in certain helix parameters of the X-ray structures are reproduced in the chemical shift derived structures. This structure derivation strategy potentially provides new directions for fast high-resolution NMR structure determination of nucleic acids, e.g. the method opens the way for fast structure determination of larger nucleic acids and of folding intermediates or low populated states for which chemical shifts could be determined

6.2 Introduction

DNAs and RNAs are central molecules of life. DNA is the carrier of genetic information, while RNAs function either as transporter of this information or regulate various cellular processes. These diverse roles require, sometimes subtle, variations in three-dimensional molecular structure and interactions, including their helical segments(1-14). Chemical shifts are highly sensitive probes of molecular structure(15-19), and in many cases, the only NMR parameters that can be obtained with any degree of completeness on a given state of a molecule. In structural biology, chemical shifts have been used to derive secondary structure of proteins(20-23) and for nucleic acids to validate their three-dimensional structure(24-28).

For proteins, several methods that employ chemical shifts as restraints for three-dimensional structure determination have been proposed and demonstrated(15; 24; 29-32). We recently presented a method(13) for the derivation of nucleic acid helix structures using non-exchangeable proton chemical shifts as sole source of experimental restraints. The method was demonstrated and validated by means of simulated data and shown to work in practice using experimental chemical shifts by the derivation of two published NMR RNA sequences. It was shown that the helix structure can be derived with a high degree of precision and accuracy. The method derives nucleic acid structures *de novo* from the proton chemical shifts using iterative singular value decomposition describing the helix structure in terms of helix parameters and building atomic resolution structure from it. To correct for small errors in the sugar-phosphate backbone of the converged chemical shift derived structures, backbone regularization is performed by means of a brief restrained molecular dynamics run keeping the aromatic planes fixed in Cartesian space. In this way, the global structure and thus helix parameters remain unchanged during backbone regularization. We have called the method CHEOPS (**che**mical shift *de nov**o*** structure derivation **p**rotocol employing **s**ingular value decomposition). At present, CHEOPS is suited for derivation of nucleic acids helix structures, which may or may not contain non-canonical base pairs.

The aim of the present paper is to show that with this method high-quality DNA helix structures can be derived with high accuracy and precision using experimental proton chemical shifts as main restraints. For this, the Dickerson dodecamer DNA helix structure was selected, because a good comparison with independently derived high-quality structures can be made. Its structure was independently determined with high-quality in solution by NMR using a large number of NOE and RDC restraints(33; 34) and with high-resolution in crystalline form by X-ray diffraction in different conditions(35; 36). In this way, the correctness of the DNA B-helix structure derived with experimental proton chemical shifts can be established in a reliable way. Finally, the chemical shift derived DNA structure was validated by comparing the experimental NOE and RDC data with those calculated from the chemical shift derived structures.

6.3 Results and discussion

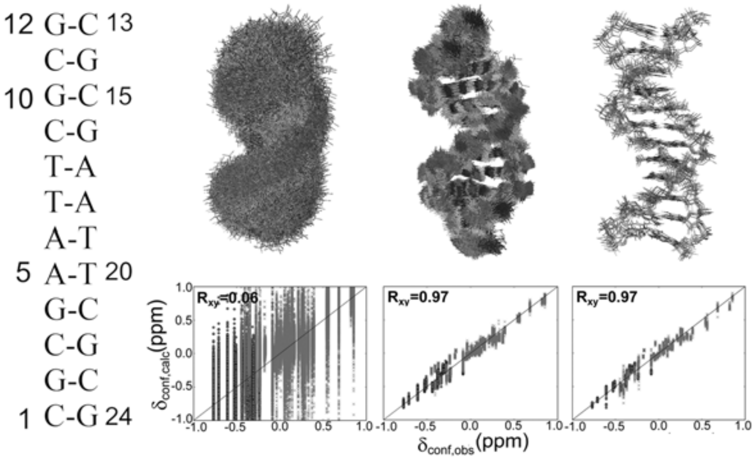


Figure 6.1: Structure derivation of the Dickerson dodecamer from proton chemical shifts (A) Sequence and secondary structure of the Dickerson dodecamer. (B) Top panel from left to right: the ensemble of 500 random starting structures, the 313 converged SVD structures, and the 8 final chemical shift structures. Bottom panel from left to right: the corresponding correlation plots of the calculated versus observed conformational proton chemical shifts, $\delta_{conf,\ calc}$ versus $\delta_{conf,\ obs}$. The $\delta_{conf,\ calc/obs}$ of the base and sugar protons are indicated as filled circles and crosses, respectively. R_{xy} stands for the Pearson correlation coefficient

The aim is to establish how precise and accurate DNA helix structures can be derived with CHEOPS using proton chemical shifts as experimental restraints. Force fields do not affect the CHEOPS derived structures, in contrast to NMR structures derived with restrained molecular dynamics that employ NOE, dihedral angle, and/or RDC restraints. Nor does homology information affect the CHEOPS structures, as is the case for protein structures derived from chemical shifts. To evaluate the performance of CHEOPS in deriving DNA helices, the structure of the Dickerson dodecamer (33-36), sequence in Figure 6.1 (left), was derived *de novo* from all 156 available proton chemical shifts (33; 37). The quality of these CHEOPS structures can be assessed by comparison with two independently derived high-resolution X-ray structures (355D and 428D, (36; 38)) and with the high-quality NMR structure, 1NAJ, which was derived with a large number of NMR restraints (33). In the following, we describe and discuss the CHEOPS structure derivation and subsequently consider how well the derived structures compare with the independently established high quality NMR and X-ray structures.

The core of the structure calculations in CHEOPS consists of an iterative SVD protocol (for details we refer to Methods and Supplementary Materials). The protocol starts here from a set of 500 random structures, built from helix parameters, uniformly distributed over ± $3\sigma_o$ around the standard B-helix values, thereby covering 99% of B-helix space (11). The set of random structures circumvent potential bias in the subsequent iterative SVD structure optimization and thus lead to accurate structures (13). The ensemble of starting structures is shown in Figure 6.1 (left set of structures) and has a pairwise RMSD of 3.6 Å (Table 6.1).

Table 6.1. Statistics of the structure derivation of the Dickerson dodecamer from experimental proton chemical shifts[$1].

	CHEOPS, Start (500 struct.)[$2]	Converged SVD (313 struct.)[$3]	Final accepted (8 struct.)[$4]	1NAJ (5 struct.)[$5]
RMSD$_{\text{chemshift}}$(all) [ppm][$6]	1.2 ± 3.8	0.079 ± 0.002	0.085 ± 0.004	0.105 ± 0.004
RMSD$_{\text{pairwise}}$ (all) [Å][$7]	3.6	1.1	1.0	0.13
RMSD$_{\text{avgto1NAJ}}$ (all/base) [Å][$8]		1.3/1.0	1.3/1.1	
RMSD$_{\text{avgto355D}}$ (all/base) [Å]		1.4/1.0	1.2/1.0	
RMSD$_{\text{avgto428D}}$ (all/base) [Å]		1.3/1.0	1.0/1.0	

[$1]For the structure calculations the CHEOPS parameter settings are given in Methods. [$2]Number of starting structures generated in CHEOPS protocol. [$3]Number of converged SVD structures. [$4]Number of final accepted structures. During backbone regularization 232 loose standard B-helix distance restraints (range 1.5 – 3.0 Å) and 330 loose standard B-helix dihedral angle restraints (range 20 – 30 °) were employed as described in Supplementary Material 'backbone regularization'; no RDCs were employed. [&5]Number of deposited high-quality NMR structures (1NAJ). [$6]RMSD of the proton chemical shifts of the ensemble of structures with respect to the observed proton chemical shifts. [$7] Pairwise RMSD of the ensemble of structures. [$8]Comparison is made between the final accepted and converged SVD structures and the deposited high-quality NMR structure (1NAJ; 10 mM Na^+/50mM K^+) as well as the two deposited high-resolution X-ray structures (355D, Na^+/Mg^{2+}-counterions in crystal; 428D, K^+/Mg^{2+}-counterions). Note that the chemical shifts were measured in buffer of 10 mM Na^+ (37). RMSD$_{\text{av21NAJ/355D/428D}}$ stands for the (average) RMSD of the average final accepted structure (or average converged SVD structure) to 1NAJ/355D/428D. The average final accepted structure (or average converged SVD structure) is built via 3DNA from the average helix parameters of the ensemble of final accepted structures (or converged SVD structures). Note finally that RMSD$_{\text{avg21NAJ}}$(base) [Å] changed from 1.0Å for the converged SVD structures before regularization to 1.1Å for the final accepted structures despite the fact that base planes have been fixed in Cartesian space during regularization. This difference is caused by the fact that a smaller number of structures were selected as acceptable

Their chemical shift RMSD from experimental values, $\mathrm{RMSD}_{\mathrm{chemshift}}$, is 1.2 ppm with a spread of 3.8 ppm (Table 6.1). In the set of starting structures, the correlation between observed and calculated chemical shifts is essentially absent, as indicated also by low Pearson correlation coefficient of 0.06 (Figure 6.1). Subsequently, the iterative SVD starts and each of the 500 starting structures is optimized in a number cycles until all chemical shifts fall within the user set chemical shift threshold of the observed chemical shifts. In each cycle new helix parameters are derived using SVD by minimizing the difference between experimental and calculated chemical shifts. Based on the new helix parameters 3DNA builds a new atomic model, which subsequently is used to calculate chemical shift using NUCHEMICS. A new cycle can then start. The protocol continues until all shifts fall within the user set chemical shift threshold. Here we find that out of the 500 starting structures, 313 converged. These converged SVD structures are shown in Figure 6.1, middle set. Their $\mathrm{RMSD}_{\mathrm{pairwise}}$ equals 1.1 Å (Table 6.1), a significant improvement over the 3.6 Å of the set of starting structures. Moreover, a tight correlation is obtained between experimental and calculated chemical shifts with a Pearson coefficient of 0.97 (Figure 6.1). Finally, the average $\mathrm{RMSD}_{\mathrm{chemshift}}$ of 0.08 ppm is quite small (24; 25). Note that the converged SVD structures are least-squares solutions, because 13 chemical shifts per basepair are employed which is larger than the number 12 of helix parameters that define the conformation of a basepair and basepair step in a helix. The chemical shift threshold for convergence is set in CHEOPS such that a relatively large number of the starting structures converge to allow for acceptable statistics. A threshold value of 0.23 ppm, corresponding to 3 $\mathrm{RMSD}_{\mathrm{chemshift}}$ and thus covering ca. 99 % of chemical shift space, was needed to achieve a significant percentage of converged structures (313 out of 500, Table 6.1, see above). Interestingly, the sugar proton shifts, but not base proton shifts, showed the largest deviations after convergence. Their deviations can be attributed most likely to non-optimal sugar conformation and/or orientation, which can be resolved via backbone regularization (see below). The resulting set of converged SVD structures was backbone regularized to correct for small errors in the conformation of the sugar rings and backbones resulting from 3DNA building of atomic models from helix parameters.

For the backbone regularization (see Methods and Supplementary Material ‘backbone regularization’), a restrained molecular dynamics run is performed in Xplor (39), while keeping the aromatic planes fixed in Cartesian space. In this way, the global structure and thus helix parameters remain unchanged during backbone regularization. In addition, loose standard B-helix backbone sequential NOE restraints and sugar-backbone torsion angles restraints (legend Table 6.1) are employed to stimulate correct formation of the backbone. The loose restraints allow for sufficient variation within the normal B-helix context. Moreover, such backbone sequential NOEs will usually be available when chemical shifts have been assigned. Small adjustments in sugar and backbone conformations may thus occur upon backbone regularization. The resulting regularized structures that pass the final quality check are called final accepted structures (Methods, Supplementary Material). Upon backbone regularization, a relatively small further decrease in the $\mathrm{RMSD}_{\mathrm{pairwise}}$ is seen (from 1.1 Å to 1.0 Å; Table 6.1,Tables S6.2/S6.3a), while the $\mathrm{RMSD}_{\mathrm{chemshift}}$ increases somewhat (here from 0.08 to 0.09 ppm; Table 6.1, Table S6.3a). This increase is mostly caused by changes in H1' and H2' proton shifts, resulting from slight reorientations in the corresponding C-H bonds (< 2°) during backbone regularization. The common ribose unit for N- and S-type puckering used in Xplor does not exactly match the individual N- and S-type ribose units used in 3DNA. This aspect contributes to C-H bond reorientation in the sugar and thus to the slight changes in proton chemical shifts. The impact of backbone regularization on the $\mathrm{RMSD}_{\mathrm{chemshift}}$ illustrates again the high sensitivity of the chemical shift to changes in structure. We also performed backbone regularizations with the published NOEs instead of the standard NOEs or in addition to the published NOEs published RDCs (Table S6.3a/b). The outcome in terms of RMSDs and violations is in all cases quite similar (Supplementary Material). The bundle of final accepted structures is shown in Figure 6.1 (right set). When progressing from the start ensemble via the set of converged SVD structures to the ensemble of final accepted structures the decrease in width of the bundle is evident ($\mathrm{RMSD}_{\mathrm{pairwise}}$ from ~4 to ~1.0 Å, Table 6.1). The correlation between calculated and observed conformational chemical shifts of these structures is similar to the converged SVD structures. In the final accepted structures, no NOE violations > 1.5 Å with respect to the experimental NOE distance restraints were seen; this includes the (base) proton distances that are fixed during backbone regularization (see below). In addition, no dihedral angle violations > 30° are seen. In conclusion precise structures are obtained based on proton chemical shifts and without force field terms affecting the helix parameters.

To validate these final accepted structures we compare them with the published set of high-quality NMR structures and with the two high-resolution X-ray structures. We first consider the structure RMSDs. The average final accepted structure (and/or average converged SVD structure) was built using 3DNA based on the average helix parameters of the entire ensemble and the RMSD towards 1NAJ, 355D and 428D was calculated, $\mathrm{RMSD}_{\mathrm{av21NAJ/355D/428D}}$ (Table 6.1). It can be gleaned from the $\mathrm{RMSD}_{\mathrm{av21NAJ/355D/428D}}$ in Table 6.1, that the average final accepted structure is quite close to 1NAJ and the two X-ray structures, although it is closest to the X-ray structure 428D. The similarity of the final accepted structures and the high-quality NMR structure 1NAJ is also evident from the $\mathrm{RMSD}_{\mathrm{chemshift}}$; $\mathrm{RMSD}_{\mathrm{chemshift}}$ of the final accepted structures is 0.085 ppm, slightly better than the 0.105 ppm of 1NAJ. We further find that the $\mathrm{RMSD}_{\mathrm{RDC}}$ of the final accepted structures is ~4 Hz (Table S6.3a/b) when RDCs are included and ~7 Hz when RDCs are not included during backbone regularization. These values are comparable to the 4.4 Hz $\mathrm{RMSD}_{\mathrm{RDC}}$ of the published NMR structures 1NAJ. The slight decrease in $\mathrm{RMSD}_{\mathrm{RDC}}$, upon inclusion of RDCs in the backbone regularization, results from small adjustments in sugar orientation (and conformation) due to sugar RDCs. In conclusion, the ensemble of final accepted structures and the deposited NMR structure 1NAJ appear similar considering structure characteristics like $\mathrm{RMSD}_{\mathrm{RDC}}$, and NOE/dihedral angle violations. A similar conclusion can be drawn when considering structure RMSDs. The structure RMSDs also show that the final accepted structures are close to the two X-ray structures. We can further conclude that the different types of backbone regularization do not influence these outcomes. A more in-depth comparison of the structures can be obtained from their helix parameters given below.

We will now specifically compare the average helix parameters of the ensemble of final accepted structures with those of the ensemble of deposited NMR structures (1NAJ) and those of the high-resolution X-ray structures (355D/428D). We first note that the RMSDs of the helix parameters ($\sigma_{\mathrm{h,\,cs}}$), which equal ca 3σ_o for the 500 starting structures, reduces to ca 0.5 to 1.1σ_o for helix parameters in the ensemble of final accepted structures (note again that backbone regularization does not affect helix parameters). Overall comparison shows further that the helix parameter of the final accepted (as well as of converged SVD) structures with respect to 1NAJ is 1.2σ_o on average and with respect to 355D/428D 1.0/1.1σ_o on average; in

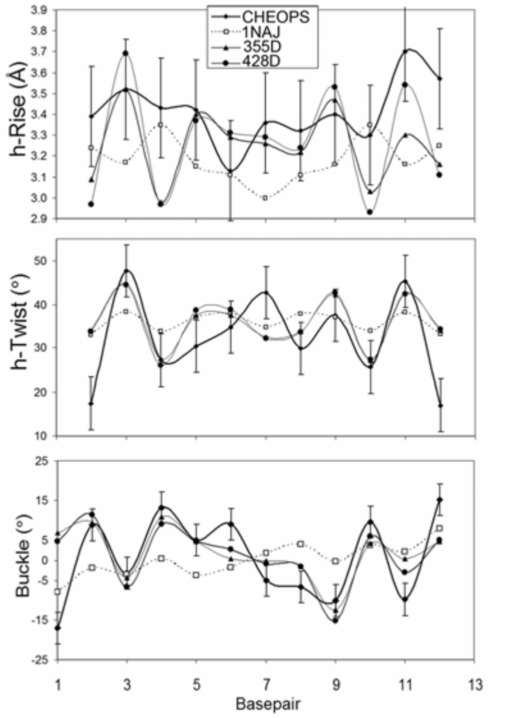


Figure 6.2: Helical rise (h-Rise, top), helical twist (h-Twist, middle), and Buckle (bottom) of the Dickerson dodecamer in the final accepted structures derived using CHEOPS compared with the corresponding parameters in the two deposited X-ray structures and in the deposited high-resolution NMR structure. The average h-Rise (top), h-Twist (middle), and Buckle (bottom) values of the ensemble of final accepted structures are represented as filled diamonds connected with a thick solid line; their RMSDs are given by the error bar. The open squares connected with dashed line represent the average h-Rise, h-Twist, and Buckle of the ensemble of 5-deposited NMR structures (1NAJ (33); Na^+/K^+-counterions; 16mM Na^+, en 50mM K^+). The filled triangles connected with a thin line represent the h-Rise, h-Twist, and Buckle of the high-resolution X-ray structure 355D (36) (Na^+/Mg^{2+}-counterions in crystal; 31 mM Na^+, 10 mM Mg^{2+}, and 38 mM spermine). The filled circles represent these parameters for the high-resolution X-ray structure 428D (38) (K^+/Mg^{2+}-counterions in crystal; 19mM K^+, 10mM Mg^{2+}, and 8.9 mM spermine). The ^1H chemical shifts used in CHEOPS were assigned from NMR spectra of samples recorded in 10 mM Na^+ (37)

this comparison, end-effects were excluded by omitting basepairs C1:G24, G2:C23, C11:G14, and G12:C13.

Thus, the final accepted structures are somewhat closer to the X-ray structures than to the NMR structure. We further investigate the degree similarity by the examples of helical rise, helical twist, and buckle (Figure 6.2). A pronounced sinusoidal variation in helical rise (h-rise) in the X-ray structures (355D/428D, Figure 6.2A) is visible. These fluctuations are strongest for the terminal three base pairs. Interestingly, in the average h-rise of the ensemble of final accepted structures these variations along the sequence are also seen but with lower amplitude at the terminal residues (Figure 6.2A). Similar observations can be made for the helical twist (Figure 6.2B) and buckle (Figure 6.2C). The variation in these parameters seen in the X-ray structures comes back in the final accepted structure. This hints to the potential of being able to determine variations in (some) helix parameters within the B-helix allowed domains from chemical shift defined DNA helix structures, although one should be aware of the similar magnitudes of helix parameter variation and $\sigma_{h,\,cs}$ of the final accepted structures. A further interesting observation is that the average h-rise of the final accepted structures is quite close to that of the two X-ray structures, in particular when excluding end-residues, where end-fraying in solution is likely to induce differences with the crystal structure(40; 41). The average h-rise of the ensemble of 1NAJ NMR structures is seen to be slightly smaller (excluding end-effects) than the h-rise of either the final accepted or the X-ray structures. Force field effects cause these smaller values as demonstrated by Kuszewski et al.(42). The 1NAJ structures were derived to unusual high quality thanks to the large number of heteronuclear RDC restraints employed beyond the NOE and dihedral restraints. However, heteronuclear RDC do not contain translational information, leaving translational parameters like h-rise somewhat undetermined. Consequently, force-field terms may still affect translational parameters like h-rise. The proton chemical shifts do contain both translational and orientational information via the ring current and magnetic anisotropy terms ($\delta_{conf,rc/ma} \propto 1/r^3$). We find that h-rise is determined by the proton chemical shifts with a good precision, $\sigma_{h,cs,h\text{-}rise} \approx 1\ \sigma_{o,h\text{-}rise}$, where $\sigma_{o,h\text{-}rise} \approx 0.28$ Å (Table S6.4). Moreover, the parameterization of ring current and magnetic anisotropy equations is good enough to obtain correct h-rise values as evident from the comparison in Figure 6.1 of the h-rise of the final accepted and X-ray structures. Finally, it is important to stress again that the chemical shift derived structures presented here do not depend on force fields terms, whereas such terms do play a role in NMR structures derived from NOE, dihedral angle, and/or RDC restraints.

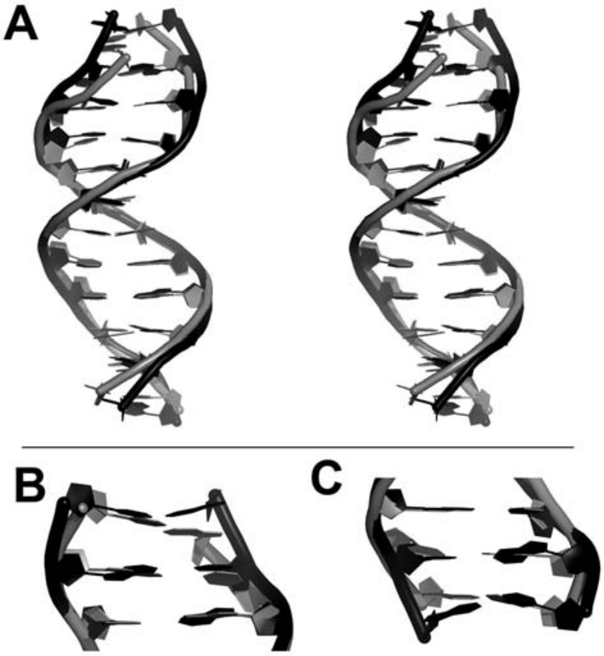


Figure 6.3: Overlay of the Dickerson dodecamer X-ray crystal structure with K^+ counterions (PDB-entry 428D (38); grey) and the average of the final accepted CHEOPS structure (black). The overlay illustrates the close correspondence between these structures as well as differences observed at the helix ends; the latter are most likely caused by end-fraying effects seen in solution(40; 41) but not in the crystal. A. Cross-eyed stereo side view of the aligned structures. High structural similarity is evident except for some end-effects. Note the unwinding (smaller helical twist, Figure 6.2) seen for the ultimate and penultimate basepair steps of the solution (CHEOPS) structure compared to the crystal structure. Also the helix rise is slightly larger for the ultimate basepairs of the solution (CHEOPS) structure. B. Illustration of the difference in buckle in the ultimate basepair 12-13 (Figure 6.2) of the solution (CHEOPS) and crystal structures (the helices are aligned on residues 10 and 15). C. Illustration of the difference in buckle in the ultimate basepair 1-24 (Figure 6.2) of the solution (CHEOPS) and crystal structures (the helices are aligned on residues 3 and 22).

It is interesting to note that the largest deviations between the helix parameters of the final accepted structures and the X-ray structures are located near the end of the helices. These largest deviations are found in the helix parameters h-rise, h-twist, propeller twist, and buckle as evident from Figure 6.2 and Table S6.4 (bold). The deviations in the ultimate residues, likely caused by end-fraying effects(40; 41), are illustrated in Figure 6.3 for h-twist, h-twist and buckle. As evident from Figure 6.2A (and Figure 6.3A), the h-rise of the X-ray structures and the final accepted structures is quite similar and only at the ultimate basepairs deviations are seen. In Figure 6.3A it can be seen that the h-rise of the ultimate basepair step is enlarged so that the ultimate base pairs no longer overlap. The helical twist shows a similar picture. As evident from Figure 6.2B the h-twist is larger for the ultimate base pair step in the final accepted structures as compared to the X-ray. In Figure 6.3 it is visualized that the helix ends are indeed slightly extra unwound. Finally, Figure 6.3B and 6.3C illustrate the increased buckle, evident in the ultimate basepairs. That these differences are seen mostly in the ultimate basepairs or helix ends suggests that they may be due to the difference between solution and crystal, i.e. are end-fraying effects usually seen under solution conditions(40; 41).

We finally note that the CHEOPS structure calculations were carried out in helix parameter space and not in atomic coordinate space. The reduced space (12 per base pair) spanned by the helix parameters allows one to focus on the essential conformational features rather than having to consider all (partly irrelevant) degrees of freedom, as would be the case with when nucleic acid conformation is described in terms of conventional atomic coordinates or torsion angles. As a result of the reduced helix parameter space smooth linear relation between helix parameter and chemical shifts are obtained. As a consequence the iterative SVD shows good convergence characteristics. Restrained molecular dynamics calculations in Cartesian space using chemical shifts as restraints are hampered by convergence problems in our experience (data not shown), most likely due to trapping in local minima.

6.4 Conclusion

We have shown that DNA helix structures can be derived based on experimental non-exchangeable proton chemical shifts as the main source of experimental information. The helices can be derived with a high degree of accuracy and precision, so that subtle variations within the B-helix domain can be established experimentally. Furthermore, the final accepted structures are not affected by force field terms like NMR structures based on NOE, dihedral angle, and RDC restraints(42). Also, no use is made of sequence homology. Thus the conformation of the final chemical shift derived structures is essentially based on chemical shift information. An extension to use RDCs and/or other restraints together with the proton chemical shifts in the iterative SVD protocol is underway. The iterative SVD protocol is now limited to nucleic acid helix regions, which may contain non-canonical base pairs. Inclusion of non-helical structure elements such as loops, bulges, and mismatches in the iterative SVD protocol is in principal possible but requires additional programming. However, it is already possible to combine the iterative SVD protocol for helical regions with restrained molecular dynamics. In this way, full structure calculation can be performed of DNAs with a stem-loop conformation or containing internal loops or bulges employing proton chemical shifts as restraints. Note CHEOPS allows calculations of structures that contain residues with N-type and/or S-type sugars as predefined by the user. Further, we note that because chemical shifts are sensitive to the dynamics on the nano- to microsecond time scale, the chemical-shift restraints can be treated as ensemble averages as described for other NMR observables (43-45). This approach should enable a description of the structural and dynamical properties of specific DNAs under a variety of conditions. It has been shown that chemical shifts can be determined of RNA folding intermediates from real-time NMR at second resolution and of low-populated states from NMR relaxation data (46-49). The ability to define detailed structures from chemical shifts as described here could be crucial in addressing the structural challenges associated with such systems and hence play an increasingly important and unique role in structural and molecular biology. Finally, we note that a direct potential application of CHEOPS is fast structure derivation of larger nucleic acids.

6.5 Methods

Deposited DNA structures and programs employed

The Dickerson dodecamer structures were retrieved from the PDB (http://www.pdb.org) and its matching proton chemical shifts from the BMRB (http: //www. bmrb. wisc.edu): 1NAJ (BMRB-entry: 5671), 355D (Na^+ counterions), and 428D (K^+ counterions). The nucleic acid structure calculations from proton chemical shifts were carried out using the program CHEOPS that we recently developed for this purpose (13) (see below). CHEOPS makes use of NUCHEMICS (24; 25) to calculate proton chemical shifts from structure and 3DNA (9) to derive helix parameters from a given structure and to build all-atom models given helix parameters. CHEOPS makes, in addition, use of XPLOR to regularize as a final step the sugar-phosphate backbone by means of a brief restrained molecular dynamics run (see below).

Nucleic acid structure calculation from experimental proton chemical shifts, CHEOPS
For the nucleic acid structure calculation from proton chemical shifts we employed the recently developed python program CHEOPS (13). Briefly, the core of CHEOPS is an iterative singular value decomposition (SVD) procedure that in a number of cycles optimizes each of the random B-helix starting structures until all calculated chemical shifts fall within the user set chemical shift threshold from the observed chemical shifts. The converged structures are called converged SVD structures. Usually, the helix parameters shear, stretch, and opening are fixed on predefined values (e.g. standard B-helix values). These parameters only weakly depend on chemical shift (13) and leaving them free to adjust would allow for inclusion unrealistic structures, slowing down the structure derivation process. Moreover, hydrogen bond restraints could be established based on imino proton resonance positions(50). The structures are described in terms of helix parameters and the program 3DNA v1.5 (9) is used to build all-atom models from these helix parameters; NUCHEMICS (24; 25) calculates the proton chemical shifts from these all-atom models. To correct the converged SVD structures for small errors in the sugar-phosphate backbone, backbone regularization is performed (Supplementary Material ‘backbone regularization’). The backbone regularization consists of a restrained molecular dynamics run with the bases kept fixed in Cartesian space, so that the global structure and helix parameters remain unchanged. The default option is to apply loose standard B-helix distance and dihedral angle restraints (51), but other restraint sets may be chosen as well. The resulting structures are quality checked in terms of adherence to holonomic restraints and applied experimental restraints during backbone regularization (e.g. NOEs, dihedral angles, RDC) and absence of hydrogen bond violations and steric clashes (checked via NUCHECK (52)). Structures that satisfy all these criteria are referred to as the final chemical shift structures or final accepted structures.

With respect to the CHEOPS protocol previously described (13), the following changes were made. 1) The input chemical shifts are here the published experimental proton chemical shifts (37). All available 156 proton chemical shifts (37) were employed as restraints. 2) The structure calculations started from a set of 500 random starting helix structures, generated within B-helix space with a variation of $\pm 3\sigma_o$ in the helix parameters around the standard B-helix values. Here, σ_o is the standard deviation of the helix parameters of a set of high-resolution X-ray B-DNA helix structures (see Olson et al.(11), Table S6.1a-top row). Consequently, $3\sigma_o$ is a measure of the allowed B-helix space and $\pm 3\sigma_o$ covers 99 % of this space. 3) For the structure calculations the following settings were employed: the chemical shift tolerance was set to 0.23 ppm (see text); the λ-value is set to 4.25; this value was found to be the optimal for providing stable SVD solutions given the error in the experimental chemical shifts of a few percent (measurement error plus estimated back-calculation error; see also van der Werf et al. (13)); the scaling factor is set to 0.25 and the maximum number of iterative cycles was set to 150.

Analysis of the chemical shift derived structures (final accepted and/or converged SVD)

The helix parameters of the set of final accepted structures (and thus of the converged SVD structures) were determined using 3DNA (9). For comparison, the helix parameters of the deposited NMR structures and of two X-ray structures (33; 36; 38) were also calculated via 3DNA. The average pairwise RMSD ($\mathrm{RMSD}_{\mathrm{pairwise}}$) of the atomic coordinates of the ensemble of final accepted structures (and or converged SVD structure) was calculated; $\mathrm{RMSD}_{\mathrm{pairwise}}$ is a measure of the precision of the protocol. In addition, the average accepted structure (and/or converged SVD structure) was built using 3DNA based on the average helix parameters of the entire ensemble. The RMSD of this average structure to 1NAJ, 355D, or 428D, $\mathrm{RMSD}_{\mathrm{av21NAJ/355D/428D}}$, is a measure of the accuracy of the protocol.

Acknowledgments. The 6th framework program of the EU, project FSG-V-RNA, is acknowledged for funding.

6.6 References

1. Kosikov KM, Gorin A, Zhurkin VB, Olson WK. 1999. DNA Stretching and Compression: Large-scale Simulations of Double Helical Structures. *J Mol Biol* 289:1301-26
2. Lu X, Shakked Z, Olson WK. 2000. A-form Conformational Motifs in Ligand-bound DNA Structures. *J Mol Biol* 300:819-40
3. Lu X, Olson WK. 1999. Resolving the Discrepancies Among Nucleic Acid Conformational Analyses. *J Mol Biol* 285:1563-75
4. Saenger W. 1988. *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. 556 pp.
5. Nelson HCM, Finch JT, Luisi BF, Klug A. 1987. The structure of an oligo(dA)Â·oligo(dT) tract and its biological implications. *Nature* 330:221-6
6. Babcock MS, Olson WK. 1994. The Effect of Mathematics and Coordinate System on Comparability and "Dependencies" of Nucleic Acid Structure Parameters. *J Mol Biol* 237:98-124
7. Babcock MS, Pednault EPD, Olson WK. 1994. Nucleic Acid Structure Analysis: Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable Between Structures. *J Mol Biol* 237:125-56
8. El Hassan MA, Calladine CR. 1995. The Assessment of the Geometry of Dinucleotide Steps in Double-Helical DNA; a New Local Calculation Scheme. *J Mol Biol* 251:648-64
9. Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108-21
10. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, et al. 1989. Definitions and Nomenclature of Nucleic Acid Structure Parameters. *J Mol Biol* 205:787-91
11. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37
12. Gorin AA, Zhurkin VB, Olson WK. 1995. B-DNA Twisting Correlates with Base-pair Morphology. *J Mol Biol* 247:34-48
13. Werf RMvd, Tessari M, Wijmenga SS. 2010. Nucleic Acid Helix Structure Determination from NMR Proton Chemical Shifts. *Submitted*
14. Musselman C, Pitt SW, Gulati K, Foster LL, Andricioaei I, Al-Hashimi HM. 2006. Impact of static and dynamic A-form heterogeneity on the determination of RNA global structure dynamics using NMR residual dipolar couplings. *J Biomol NMR* 36:235-49
15. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. 2007. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615-20
16. Wutrich K. 1986. *NMR of Proteins and Nucleic Acids*. New York: John Wiley & Sons. 291 pp.
17. Cornilescu G, Delagio F, Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289-302

18. Wishart DS, Case DA. 2002. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* 338:3-34

19. Sanders JKM, Hunter BK. 1993. *Modern NMR Spectroscopy, a guide for chemists*. Oxford: Oxford University Press. 330 pp.

20. Wishart DS, Sykes BD. 1994. The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J Biomol NMR* 4:171-80

21. Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31:1647-51

22. Mielke SP, Krishnan VV. 2009. Characterization of protein secondary structure from NMR chemical shifts. *Prog NMR Spectrosc* 54:141-65

23. Williamson MP, Asakura T. 1997. Protein chemical shifts. *Methods Mol. Biol.* 60:53-9

24. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. 2001. Prediction of Proton Chemical Shifts in RNA. Their use in Structure refinement and validation. *J Biomol NMR* 21:11-29

25. Wijmenga SS, Kruithof M, Hilbers CW. 1997. Analysis of ^{1}H chemical shifts in DNA: Assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* 10:337-50

26. Santini GPH, Cognet JAH, Xu D, Singarapu KK, Penhoat CHd. 2009. Nucleic Acid Folding Determined by Mesoscale Modeling and NMR Spectroscopy: Solution Structure of d(GC$\underline{\text{GAAA}}$GC). *J.Phys.Chem.B* 113:6881-93

27. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on Duck, Heron and Human HBV encapsidation signals. *Nucleic Acids Res* 35:2800-11

28. Lam SL, Chi LM. 2010. Use of chemical shifts for structural studies of nucleic acids. *Prog NMR Spectrosc* 56:289-310

29. Fars C, Amata I, Carlomagno T. 2007. C-Detection in RNA Bases: Revealing Structure–Chemical Shift Relationships. *J Am Chem Soc* 129:15814-23

30. Shen Y, Lange O, Delagio F, Rossi P, Aramini JM, et al. 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685-90

31. Shen Y, Vernon R, Baker D, Bax A. 2009. De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* 43:63-78

32. Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G. 2008. CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496-W502

33. Wu Z, Delagio F, Tjandra N, Zhurkin VB, Bax A. 2003. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ^{31}P chemical shift anisotropy. *J Biomol NMR* 26:297-315

34. Tjandra N, Tate S, Ono A, Kainosho M, Bax A. 2000. The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase. *J Am Chem Soc* 122:6190-200

35. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, et al. 1981. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc Natl Acad Sci USA* 78:2179-83

36. Shui X, McFail-Isom L, Hu GG, Williams LD. 1998. The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* 37:8341-55

37. Hare DR, Wemmer DE, Chou S, Drobny G, Reid BR. 1983. Assignment of the Non-exchangeable Proton Resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) Using Two-dimensional Nuclear Magnetic Resonance Methods. *J Mol Biol* 171:319-36

38. Shui X, Sines CC, McFail-Isom L, Veer Dvd, Williams LD. 1998. Structure of the Potassium Form of CGCGAATTCGCG: DNA Deformation by Electrostatic Collapse around Inorganic Cations. *Biochemistry* 37:16877-87

39. Brunger AT. 1992. *X-PLOR version 3.1 A system for X-ray Crystallography and NMR*. New Haven: Yale University Press

40. Andreatta D, Sen S, Lustres JLP, Kovalenko SA, Ernsting NP, et al. 2006. Ultrafast Dynamics in DNA: “Fraying” at the End of the Helix. *J Am Chem Soc* 128:6885-92

41. Patel DJ, Hilbers CW. 1975. Proton Nuclear Magnetic Resonance Investigations of Fraying in Double-Stranded d-ApTpGpCpApT in H_20 Solution. *Biochemistry* 14:2651-6

42. Kuszewski JJ, Schwieters CD, Clore GM. 2001. Improving the Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base−Base Positional Interactions. *J Am Chem Soc* 123:3903-18

43. Vendruscolo M, Dobson CM. 2005. Towards complete descriptions of the free-energy landscapes of proteins. *Phil.Trans.R.Soc.A* 363:433-52

44. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. 2005. Simultaneous determination of protein structure and dynamics. *Nature* 433:128-32

45. Clore GM, Schwieters CD. 2006. Concordance of Residual Dipolar Couplings, Backbone Order Parameters and Crystallographic B-factors for a Small α/β Protein: A Unified Picture of High Probability, Fast Atomic Motions in Proteins *J Mol Biol* 355:879-86

46. Ampt KAM, Werf RMvd, Nelissen FHT, Tessari M, Wijmenga SS. 2009. The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding. *Biochemistry* 48:10499-508

47. Korzhnev DM, Kay LE. 2007. Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Acc Chem Res* 41:442-51

48. Shajani S, Varani G. 2007. NMR Studies of Dynamics in RNA and DNA by ^{13}C Relaxation. *Biopolymers* 86:348-59

49. Hoogstraten CG, Wank JR, Pardi A. 2000. Active Site Dynamics in the Lead-Dependent Ribozyme. *Biochemistry* 39:9951-8

50. Grzesiek S, Cordier F, Jaravine V, Barfield M. 2004. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog NMR Spectrosc* 45:275-300

51. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387

52. Feng Z, Westbrook J, Berman HM. 1998. NUCheck. *Rep. NDB-407*, Rutgers University, New Brunswick

6.7 Supplementary Material

Regularization of the backbone

Backbone regularization was applied to correct for small errors in the conformation of the sugar and backbone in the chemical shift optimized structures (see Materials and Methods). The backbone regularization consists of a short restrained Molecular Dynamics (rMD) protocol which was run in Xplor(1) v3.851. During the rMD the aromatic planes are kept fixed in atomic Cartesian space, so that helix parameters remain unchanged, thereby preserving the overall structure. Standard holonomic restraints were used with bond lengths and bond angles taken from Gelbin et al. (2) and improper dihedral angles from Saenger(3). Dihedral angles in the backbone and sugar rings as well as the dihedral angle χ were loosely restrained to standard B-helix values as given by Gelbin et al. (2). Also, standard hydrogen bond distance restraints in the basepairs (4; 5) were included in the protocol. In addition, we included loose standard B-helix inter-residue sugar-sugar and sugar-aromatic distances as given by Wijmenga et al. (6). For the distance restraints a soft-square potential with force constant of 45 kcal mole^{-1} (Å)$^{-2}$ was used. For the dihedral restraints a harmonic potential was used of 150 kcal mole^{-1} (°)$^{-2}$. The boundaries of the loose standard B-helix dihedral restraints were set to ± 20° except for the backbone dihedral angles α, β, γ, and ε, which were set to ± 30°. The loose standard B-helix NOE distance restraints consist of inter-residue distances with different upper and lower boundaries as follows: $H2'_i$-$H1'_{i+1}$, $H2'_i$-$H5'_{i+1}$, $H2'_i$-$H5_{i+1}$: upper and lower boundaries of 1.0Å; $H2'_i$-$H6/8_{i+1}$: upper and lower boundaries of 1.25 Å; $H1'_i$-$H6/8_{i+1}$, and $H1'_i$-$H5'_{i+1}$: with upper boundary of 1.50 Å and lower boundary of 0.50Å. For the Dickerson dodecamer 232 loose standard B-helices distance restraints were used, complemented with 330 loose standard B-helix dihedral angle restraints.

The restrained Molecular Dynamics protocol consisted of a Powell Minimization of 20,000 steps to dissipate energy resulting of holonomic restraint violation. This was followed by 25 ps of Verlet dynamics in 100,000 steps, followed by a final Powell minimization of 10,000 steps. When RDCs are included in the calculation, the above rMD protocol is repeated with the RDC potential switched on with a force constant of 0.2 cal^2/Hz (harmonic potential). Based on adherence to employed restraints (no distance violations larger than 1.50 Å, no dihedral angle violations larger than 30°) structures were accepted.

In summary the following different backbone regularizations were carried out. 1) The backbone regularization as described above was performed, i.e. during the rMD the base planes are kept fixed and 232 standard NOE/ 330 dihedral restraints are employed but no RDC restraints. 2) Backbone regularization as described in point 1 was performed and subsequently repeated with 204 published RDC restraints included (as given in BMRB). 3) Backbone regularization as described in point 1 was performed with the 232 standard NOE restraints replaced with 146 experimental NOE restraints (as given in BMRB). 4) Backbone regularization as described in point 2 was performed with 232 standard NOE restraints replaced by 146 experimental NOE restraints. 5) Backbone regularization as described in point 4 with 22 published dihedrals (no standard dihedrals).

Table S6.1a. Helix parameters of the deposited NMR structure of the Dickerson dodecamer, 1NAJ (7)

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.21	-0.15	0.12	0.09	0.19	0.50	6.70	-11.40	5.30	0.60	3.10
C-G	0.35	0.05	-0.52	0.02	-0.36	0.03	-8.05	1.32	-17.90	0.51	1.39	1.04
G-C	-0.14	0.03	-0.34	0.03	0.35	0.03	-2.04	1.07	-14.55	1.14	0.85	0.14
C-G	0.41	0.08	-0.47	0.04	-0.05	0.11	-3.69	0.43	-14.18	1.53	0.81	1.09
G-C	-0.19	0.08	-0.39	0.02	0.01	0.04	0.32	0.35	-13.36	1.10	-1.26	0.98
A-T	-0.20	0.05	-0.27	0.05	-0.18	0.04	-3.89	0.31	-19.64	0.14	-1.13	0.76
A-T	-0.06	0.01	-0.23	0.06	-0.24	0.05	-1.81	0.83	-22.94	0.17	-2.76	0.55
T-A	0.07	0.02	-0.24	0.06	-0.24	0.05	1.76	0.77	-22.96	0.10	-2.68	0.64
T-A	0.20	0.05	-0.27	0.06	-0.19	0.05	3.97	0.47	-19.47	0.35	-1.00	0.99
C-G	0.18	0.07	-0.39	0.02	0.03	0.03	-0.34	0.19	-13.31	1.03	-1.33	0.92
G-C	-0.40	0.07	-0.46	0.03	-0.05	0.08	3.72	0.53	-14.32	1.57	0.77	0.88
C-G	0.15	0.01	-0.34	0.02	0.35	0.01	2.12	0.62	-14.59	0.85	0.77	0.21
G-C	-0.34	0.05	-0.52	0.04	-0.35	0.08	8.06	0.92	-17.86	0.56	1.37	0.98

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.05	1.28	0.02	0.87	3.29	0.21	2.10	9.20	0.00	4.30	36.50	6.60
C-G	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-1.19	0.11	-1.30	0.18	3.24	0.09	15.56	1.69	5.70	0.23	33.02	0.24
C-G	-0.66	0.19	0.53	0.07	3.17	0.05	0.36	0.27	-4.55	1.18	38.39	0.74
G-C	-1.25	0.06	0.16	0.22	3.35	0.02	11.41	1.82	-3.44	0.66	33.83	0.55
A-T	-0.98	0.08	0.54	0.10	3.15	0.02	8.60	0.19	0.73	0.46	37.06	0.15
A-T	-0.33	0.08	0.63	0.04	3.11	0.01	-0.20	0.90	0.85	0.30	37.88	0.17
T-A	-0.70	0.08	0.01	0.04	3.00	0.05	-4.38	0.91	-0.02	0.09	34.69	0.13
T-A	-0.31	0.03	-0.60	0.07	3.11	0.01	-0.08	0.86	-0.98	0.49	37.91	0.17
C-G	-0.94	0.09	-0.54	0.09	3.16	0.00	8.54	0.55	-0.52	0.23	37.10	0.24
G-C	-1.19	0.07	-0.16	0.17	3.35	0.01	11.32	1.72	3.27	0.53	33.91	0.50
C-G	-0.67	0.18	-0.57	0.04	3.16	0.05	0.45	0.27	4.62	0.92	38.12	0.42
G-C	-1.19	0.13	1.27	0.12	3.25	0.05	15.65	1.24	-5.77	0.53	33.19	0.28

Helix parameters deviating more than 2σo from standard B-helix values as defined in Olson et al. (8) are marked in bold; σo is the RMSD of the helix parameter indicating allowed B-helix conformational space as derived based on X-ray structures(9).

Table S6.1b. Helix the deposited X-ray structure of the Dickerson dodecamer with Na^+ counter ions, 355D(10)

	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
Standard	0.00	-0.15	0.09	0.50	-11.40	0.60	0.05	0.02	3.29	2.10	0.00	36.50
C-G	0.28	-0.14	0.07	6.93	-17.31	-0.61	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-0.24	-0.18	0.49	9.34	-14.30	-2.08	-1.27	-0.65	3.09	14.77	5.57	33.94
C-G	0.24	-0.17	0.16	-4.43	-5.40	0.43	1.79	-0.36	3.52	-11.95	-3.76	44.84
G-C	-0.26	-0.11	0.01	10.81	-9.45	1.01	-1.40	0.52	2.98	24.46	-2.09	27.52
A-T	-0.04	-0.11	0.01	4.72	-15.31	1.60	-0.40	0.49	3.40	2.13	2.46	37.56
A-T	0.05	-0.05	0.07	0.44	-15.00	6.23	-0.57	-0.31	3.29	0.71	0.51	37.52
T-A	-0.04	-0.12	0.17	-0.26	-16.74	3.93	-0.61	-0.03	3.26	-4.78	0.56	32.51
T-A	-0.11	-0.12	0.00	-1.56	-16.36	5.12	-0.53	0.41	3.22	-1.68	-2.89	33.80
C-G	0.21	-0.13	0.00	-12.41	-10.27	-1.22	-0.13	0.45	3.47	-2.34	-0.95	42.17
G-C	-0.11	-0.05	0.24	4.21	-9.60	3.21	0.66	-2.36	3.03	8.97	7.84	27.14
C-G	0.16	-0.13	0.21	0.28	-17.42	-1.75	1.47	1.50	3.30	-12.97	3.95	42.69
G-C	-0.24	-0.07	0.25	4.67	-4.95	-1.62	-0.69	0.42	3.16	11.53	1.17	33.96

Helix parameters deviating more than $2\sigma_o$ from standard B-helix values as defined in Olson et al.(8) are marked in bold; σ_o is the RMSD of the helix parameter indicating allowed B-helix conformational space as derived based on X-ray structures (9).

Table S6.1c. Helix the deposited X-ray structure of the Dickerson dodecamer with K^+ counter ions, 428D(11)

	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
Standard	0.00	-0.15	0.09	0.50	-11.40	0.60	0.05	0.02	3.29	2.10	0.00	36.50
C-G	0.34	-0.17	0.16	4.77	-18.36	1.29	0.00	0.00	0.00	0.00	0.00	0.00
G-C	-0.31	-0.29	0.70	11.40	-13.07	-2.07	0.04	0.07	3.06	-4.21	7.80	32.72
C-G	0.25	-0.08	0.35	-6.78	-6.74	1.27	0.51	0.60	3.85	2.86	-9.13	43.50
G-C	-0.69	-0.08	-0.03	9.07	-9.04	6.06	-0.10	0.64	3.00	3.37	10.01	23.76
A-T	0.00	-0.07	0.00	4.64	-17.23	2.94	-0.45	-0.21	3.38	-2.60	4.02	38.30
A-T	0.23	0.07	0.09	2.72	-15.51	5.38	0.10	-0.30	3.30	-0.69	-1.94	38.82
T-A	-0.01	0.15	-0.07	-0.92	-16.42	3.64	-0.01	-0.65	3.27	2.58	-1.47	31.97
T-A	-0.16	-0.10	0.03	-1.70	-15.49	3.44	-0.04	-0.40	3.23	-1.08	-0.45	33.58
C-G	0.26	-0.05	0.18	-15.26	-9.80	-0.47	-0.23	-0.34	3.52	0.56	-3.29	42.53
G-C	-0.07	0.01	0.36	6.02	-5.70	4.76	0.77	0.69	2.95	-2.69	3.77	27.01
C-G	0.11	-0.05	0.49	-3.16	-16.29	-2.90	-1.27	0.28	3.60	-4.47	-10.65	40.85
G-C	-0.05	-0.11	0.40	5.06	-8.76	-2.61	-0.20	0.23	3.12	1.92	4.58	33.91

Helix parameters deviating more than $2\sigma_o$ from standard B-helix values as defined in Olson et al. (8) are marked in bold; σ_o is the RMSD of the helix parameter indicating allowed B-helix conformational space as derived based on X-ray structures (9).

Table S6.2. Statistics of the chemical shift optimized structure of the Dickerson dodecamer (12 basepairs), based on 156 experimental chemical shifts from BMRB.

	Start (500 struct.)	Converged SVD (313 struct.)	1NAJ, submitted (5 struct.)
RMSD$_{\mathrm{CS}}$(all) [ppm]	1.2 ± 3.8	0.079 ± 0.002	0.105 ± 0.004
RMSD$_{\mathrm{CS}}$(base) [ppm]	0.4 ± 0.4	0.078 ± 0.003	0.096 ± 0.010
RMSD$_{\mathrm{CS}}$(sugar) [ppm]	1.3 ± 4.4	0.079 ± 0.002	0.108 ± 0.003
RMSD$_{\mathrm{NOE}}$ (published) [Å]			0.025
RMSD$_{\mathrm{RDC}}$ [Hz]	9.68 ± 0.72	-	4.39 ± 0.07
RMSD$_{\mathrm{pairwise}}$ (all) [Å]	3.63 ± 0.74	1.09 ± 0.29	0.13 ± 0.06
RMSD$_{\mathrm{pairwise}}$(base) [Å]	2.95 ± 0.64	0.87 ± 0.22	0.13 ± 0.06
RMSD$_{\mathrm{avgto1NAJ}}$ (all) [Å]	1.4	1.3	-
RMSD$_{\mathrm{avgto1NAJ}}$ (base) [Å]	1.0	1.0	-
RMSD$_{\mathrm{avgto355D}}$ (all) [Å]	1.4	1.4	1.03
RMSD$_{\mathrm{avgto355D}}$ (base) [Å]	1.0	1.0	0.97
RMSD$_{\mathrm{avgto428D}}$ (all) [Å]	1.4	1.3	1.0
RMSD$_{\mathrm{avgto428D}}$ (base) [Å]	1.0	1.0	1.0

Table S6.3a. Statistics of the final accepted structures of the Dickerson dodecamer (12 basepairs) based on 156 experimental chemical shifts from BMRB. During backbone regularization the standard distance restraints for B-helices were employed, i.e. a set of standard NOEs (232 restraints); this set was used together with standard B-helix dihedral angle restraints (330 restraints) as described above in ‘backbone regularization’. No RDCs were employed

1NAJ	Reg. (8 struct.) NOE-set: standard RDCs excluded
RMSD$_{CS}$(all) [ppm]	0.085 ± 0.004
RMSD$_{CS}$(base) [ppm]	0.084 ± 0.006
RMSD$_{CS}$(sugar) [ppm]	0.085 ± 0.004
RMSD$_{NOE}$ (published) [Å]	0.224 ± 0.023
RMSD$_{RDC}$ [Hz]	7.4 ± 0.3
RMSD$_{pairwise}$ (all) [Å]	1.0 ± 0.2
RMSD$_{pairwise}$(base) [Å]	0.8 ± 0.2
RMSD$_{avgto1NAJ}$ (all) [Å]	1.3
RMSD$_{avgto1NAJ}$ (base) [Å]	1.1
RMSD$_{avgto355D}$ (all) [Å]	1.2
RMSD$_{avgto355D}$ (base) [Å]	1.0
RMSD$_{avgto428D}$ (all) [Å]	1.0
RMSD$_{avgto428D}$ (base) [Å]	1.0

Note that RMSD$_{pairwise}$(base) [Å] has changed somewhat with respect to the value before regularization despite the fact that base planes have been fixed in Cartesian space. The relative base conformation did not change but a smaller number of structures was selected as acceptable and this has led to lower RMSD.

Table S6.3b. Statistics of the final accepted structures of the Dickerson dodecamer (12 basepairs) based on 156 experimental chemical shifts from BMRB. During backbone regularization the standard distance restraints for B-helices were replaced by the set of published NOEs (146 restraints); this set was used together with standard B-helix dihedral angle restraints (330 restraints) as described above in 'backbone regularization' with the one exception of the last column, where the standard set of dihedrals was replaced by the published set of ε dihedral angle restraints (22 restraints). The number of RDC restraints is 204, of which 158 are sugar C-H bond vectors, and 46 aromatic C-H and N-H bond vectors.

	Reg. (57 struct.) NOE-set: pub. RDCs excluded	Reg. (20 struct.) NOE-set: pub. Published RDCs included	Reg. (18 struct.) NOE-set: pub. Published RDCs and Published dihedrals included
$RMSD_{CS}$(all) [ppm]	0.082 ± 0.003	0.08 ± 0.002	0.078 ± 0.003
$RMSD_{CS}$(base) [ppm]	0.083 ± 0.004	0.084 ± 0.002	0.082 ± 0.003
$RMSD_{CS}$(sugar) [ppm]	0.081 ± 0.003	0.078 ± 0.003	0.077 ± 0.003
$RMSD_{NOE}$ (published) [Å]	0.256 ± 0.033	0.262 ± 0.025	0.264 ± 0.028
$RMSD_{RDC}$ [Hz]	7.04 ± 0.47	4.03 ± 0.08	4.23 ± 0.23
$RMSD_{pairwise}$ (all) [Å]	1.0 ± 0.3	0.9 ± 0.2	0.9 ± 0.2
$RMSD_{pairwise}$(base) [Å]	0.8 ± 0.2	0.7 ± 0.2	0.7 ± 0.2
$RMSD_{avgto1NAJ}$ (all) [Å]	1.3	1.3	1.4
$RMSD_{avgto1NAJ}$ (base) [Å]	1.1	1.1	1.1
$RMSD_{avgto355D}$ (all) [Å]	1.2	1.3	1.3
$RMSD_{avgto355D}$ (base) [Å]	1.1	1.1	1.1
$RMSD_{avgto428D}$ (all) [Å]	1.2	1.2	1.2
$RMSD_{avgto428D}$ (base) [Å]	1.0	1.1	1.1

Table S6.4. Helix parameters of the final accepted (chemical shift derived) B-DNA helix structures of the Dickerson dodecamer, after backbone regularization employing 232 standard NOE distances and 330 standard B-helix dihedral angle restraints.

	Shear		Stretch		Stagger		Buckle		Prop-Tw		Opening	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.00	0.21	-0.15	0.12	0.09	0.19	0.50	6.70	-11.40	5.30	0.60	3.10
C-G	0.00	0.00	-0.15	0.00	-0.06	0.24	-17.00	3.33	-18.28	7.07	0.60	0.01
G-C	0.00	0.00	-0.15	0.00	0.07	0.26	8.93	3.66	-19.70	3.97	0.60	0.01
C-G	0.00	0.00	-0.15	0.00	0.16	0.30	-3.30	5.22	-10.80	4.24	0.60	0.01
G-C	0.00	0.00	-0.15	0.00	0.26	0.27	13.18	4.48	-14.92	5.38	0.60	0.01
A-T	0.00	0.00	-0.15	0.00	0.11	0.27	5.06	2.98	-15.62	7.86	0.60	0.01
A-T	0.00	0.00	-0.15	0.00	0.16	0.31	9.05	3.85	-16.48	5.12	0.60	0.01
T-A	0.00	0.00	-0.15	0.00	-0.03	0.36	-5.10	4.40	-10.10	8.46	0.60	0.01
T-A	0.00	0.00	-0.15	0.00	0.00	0.38	-6.75	5.74	-20.01	5.60	0.60	0.00
C-G	0.00	0.00	-0.15	0.00	0.20	0.26	-10.08	3.76	-9.87	8.53	0.60	0.01
G-C	0.00	0.00	-0.15	0.00	0.13	0.29	9.61	4.74	-13.16	7.72	0.61	0.01
C-G	0.00	0.00	-0.15	0.00	0.21	0.21	-9.87	3.33	-20.74	5.16	0.60	0.01
G-C	0.00	0.00	-0.15	0.00	-0.04	0.18	15.21	4.88	-15.93	5.31	0.60	0.01

	X-disp		Y-disp		h-Rise		Incl.		Tip		h-Twist	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Standard	0.05	1.28	0.02	0.87	3.29	0.21	2.10	9.20	0.00	4.30	36.50	6.60
C-G	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G-C	0.23	1.57	0.39	1.24	3.39	0.24	9.03	7.35	0.91	5.56	17.43	1.75
C-G	-0.80	0.50	-0.10	0.53	3.52	0.24	13.77	5.78	1.79	2.44	47.68	5.29
G-C	1.38	0.82	-1.14	1.00	3.43	0.22	-3.02	5.35	0.58	4.44	27.28	3.98
A-T	-1.08	0.59	0.57	0.73	3.42	0.30	9.21	6.26	1.17	4.02	30.34	6.49
A-T	-0.17	0.50	-0.96	0.25	3.13	0.19	-3.68	6.57	3.72	4.01	34.77	5.80
T-A	-1.55	0.23	0.24	0.33	3.36	0.21	7.50	3.67	-0.12	2.16	42.63	4.64
T-A	-0.59	0.97	0.27	0.85	3.32	0.32	-1.35	5.76	-5.27	5.59	29.98	8.96
C-G	-1.41	0.47	0.66	0.94	3.40	0.24	2.57	6.38	-0.94	5.06	37.51	9.10
G-C	1.01	0.75	-0.33	1.47	3.30	0.22	-0.13	4.43	5.39	4.81	25.70	6.33
C-G	-1.13	0.65	-0.04	0.47	3.70	0.18	15.83	6.46	-0.85	3.26	45.35	4.06
G-C	-0.71	1.46	-0.40	0.69	3.57	0.23	12.00	7.61	-1.16	3.69	16.99	0.41

Helix parameters deviating more than $2.2\sigma_o$ are marked in bold; σ_o is the RMSD of the helix parameter indicating allowed B-helix conformational space as derived based on X-ray structures(9). Not that the helix parameters of the converged SVD structures and the final accepted structures are the same because during backbone regularization the bases planes are kept fixed in Cartesian space.

References

1. Brunger AT. 1992. X-PLOR version 3.1 A system for X-ray Crystallography and NMR. New Haven: Yale University Press
2. Gelbin A, Schneider B, Clowney L, Hsieh SH, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Am Chem Soc* 118:519-29
3. Saenger W. 1988. *Principles of Nucleic Acid Structure*. New York: Springer-Verlag. 556 pp.
4. Ping H, Eisenberg M. 1992. The Three-Dimensional Structure in Solution (pH 5.8) of a DNA 9-mer Duplex Containing 1 ,M-Propanodeoxyguanosine Opposite Deoxyadenosine. Restrained Molecular Dynamics and NOE-Based Refinement Calculations. *Biochemistry* 31:6518-32
5. Arnott S, Hukins DWL. 1973. Refinement of the Structure of B-DNA and Implications for the Analysis of X-ray Diffraction Data from Fibers of Biopolymers. *J Mol Biol* 81:93-105
6. Wijmenga SS, Buuren BNMv. 1998. The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Prog NMR Spectrosc* 32 287-387
7. Wu Z, Delagio F, Tjandra N, Zhurkin VB, Bax A. 2003. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ^{31}P chemical shift anisotropy. *J Biomol NMR* 26:297-315
8. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229-37
9. Werf RMvd, Tessari M, Wijmenga SS. 2010. Nucleic Acid Helix Structure Determination from NMR Proton Chemical Shifts. *Submitted*
10. Shui X, McFail-Isom L, Hu GG, Williams LD. 1998. The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* 37:8341-55
11. Shui X, Sines CC, McFail-Isom L, Veer Dvd, Williams LD. 1998. Structure of the Potassium Form of CGCGAATTCGCG: DNA Deformation by Electrostatic Collapse around Inorganic Cations. *Biochemistry* 37:16877-87

Chapter 7

Summary

Samenvatting

7 Summary

RNA may function as the conveyor of genetic information, as a catalytic molecule like in ribozymes, or as part of a mechanism to regulate gene expression, i.e. RNA-switches. To understand how RNA functions, it is important to study the molecular structure of RNA molecules. The research in this thesis described new methods to solve nucleic acid structures by NMR spectroscopy.

In the introduction, ways to describe RNA structures are given. In addition, structure determination of nucleic acids by NMR spectroscopy is described in more detail. The latter holds amongst others, a formal derivation of the equations used for deriving structure information from residual dipolar couplings and a description of how to use both spin-relaxation and residual dipolar couplings to derive information on internal dynamics of a nucleic acid in solution. In addition to the biophysical sections of the introduction, a description is given of the molecular biological mechanism of virus replication of the hepatitis B virus. More specific, the essential role of the ε RNA element in the hepatitis B virus replication is described.

In the second chapter, we described the structure and dynamics of the apical stem of the ε encapsidation signal of the hepatitis B virus of the duck. This because most molecular details on hepatitis B virus replication stem from the duck HBV strain, since in-vitro duck ε shows both binding and priming competence towards the protein P complex, whereas the in-vitro human ε only shows binding to the protein P complex. The structure has been determined by including a large set of ‘classical’ NMR restraints into structure calculation, i.e. NOE distance restraints and RDC restraints. We derived local dynamics from spin relaxation and relaxation dispersion measurements, and found that the structure shows enhanced basepair opening, that is crucial for protein binding, and thus the function of this domain of the RNA molecule.

In the third and fourth chapter, NMR studies of the ε encapsidation signal of the hepatitis B virus are described. A better understanding of the ε RNA structure and thus its function would be essential for designing medication against chronic hepatitis B virus infection. First, the assignment-strategy for NMR spectra of the primer binding loop of the ε encapsidation signal of the hepatitis B virus of the duck is given. Second, the structure and dynamics of the full ε of human and duck were determined using NOE-distance information, RDCs, and spin relaxation data. RDCs were measured from a sample with pf1 phages as aligning medium. These RDCs were used to visualize the domain motions of the helices of the full ε. It was found that both human and duck apical stems make a hinging motion of about 20°, and the primer-binding loop causes a bend in the ε structure. From further studies on dynamics, it follows that the primer synthesis site is exposed from time to time, and we suggest this paves the road for conformational capturing of the open primer synthesis in the active site of the P-protein, that is responsible for genome replication.

In the fifth and sixth chapter, the relationship between proton chemical shift and base positioning in nucleic acid helix structures is established. This led to a new calculation protocol that allows for calculation of helical nucleic acid structure based on isotropic non-exchangeable proton chemical shifts. It makes use of the fact that ring currents in aromatic rings dominate the proton chemical shifts. The protocol, called CHEOPS, optimizes the proton chemical shift values by positioning the base planes via an iterative singular value decomposition method. This singular value decomposition makes the protocol suitable for structure determination using an incomplete list of proton chemical shift values. The positions of base planes can be determined down to 0.6 Å, and orientation down to 6°. After positioning the base planes, the molecule is subjected to molecular dynamics simulation to correct for small deviations in the backbone. Structures can be determined with high precision and accuracy with a list of proton chemical shift lists containing as little as four chemical shifts per basepair. Besides a thorough analysis of the range of validity of CHEOPS, it is demonstrated on two known RNA structures, which were derived based employing only proton chemical shifts as restraints.

In the sixth chapter, the method was further demonstrated using experimental proton chemical shift values, to determine *de novo* the Dickerson dodecamer B-DNA helix structure. Our calculation method was able to reproduce the helical segments of the structure, satisfying deposited NMR restraints (such as NOE and RDCs) as well as isotropic proton chemical shift values. In addition, our NMR derived structure does show similarities with the X-ray structures. This underlines the quality of our calculation method, since the quality of X-ray structures is known to be better than the quality of ‘classical’ NMR structures.

Research from chapters two to four resulted in structure and dynamics a biological relevant viral RNA, which could hold the key for drug development against chronic HBV infection. Furthermore, the fourth chapter provides a method to employ phage induced RDCs to determine and visualize molecular dynamics of nucleic acid domains. This complemented with the results from chapters five and six, where a method was shown to calculate helix structure from proton chemical shifts, introduces a variety of new methods to determine structure and dynamics by NMR.

Prospect

In this thesis the determination of structure and dynamics of the ε element of the hepatitis B virus were described. Further we inferred on the binding of ε in the reverse-transcriptase domain of the P-protein. Further studies of the complex of P-protein and ε may be used to confirm this model in more detail. For these studies, structures of the different domains of the P-protein complex may be required, and their structures must be determined with and without bound ε RNA.

When considering the methodological aspects of this thesis, we described a method to derive helical nucleic acid structures employing only isotropic proton chemical shifts. Additional research is required to apply this method to non-helical elements, such as bulges and loops. In addition, the question rises how this method can be applied employing both proton and heteronuclear chemical shifts, i.e. ^{13}C, ^{15}N, and ^{31}P. Furthermore, additional studies are required to determine a structure with a combination of isotropic chemical shifts and other restraints, such as RDCs and NOEs.

Inclusion of internal dynamics on the isotropic chemical shifts requires further studies. When that theory has been further developed, one could derive information on internal dynamics from chemical shifts.

The proposed protocol in chapter 4 and 5 opens up a new way of structure determination. Once the method is further developed, it may provide a tool to determine even larger structure with high accuracy and precision, than is possible with NMR spectroscopy now. In addition, the method would allow for deriving structures of low populated states of nucleic acids to provide a better understanding of the role these structures play in the function of the molecule.

Samenvatting

RNA is een molecuul dat kan functioneren als boodschapper van de genetische code, als katalysator, zoals in ribozymen, of als onderdeel van een mechanisme om gen-expressie te reguleren, bijvoorbeeld RNA-switches. Om te begrijpen hoe RNA werkt, is het van belang om de structuur van deze moleculen te bestuderen. Het onderzoek in dit proefschrift beschrijft nieuwe methoden om RNA structuren op te lossen met behulp van NMR spectroscopie.

In de inleiding worden een aantal manieren gegeven om RNA structuren te beschrijven. Verder wordt er in detail beschreven, hoe structuren van nucleine zuren kunnen worden opgelost met behulp van NMR spectroscopie. In dit deel van de inleiding wordt onder meer een formele afleiding gegeven voor de vergelijkingen die het mogelijk maken, structuur en dynamica informatie uit residuele dipolaire koppelingen (RDCs) te halen. Naast de biofysische secties van de inleiding, wordt er een beschrijving gegeven van het moleculair biologische mechanisme van virus replicatie van het hepatitis B virus. Meer specifiek, de essentiele rol van het ε RNA element in de hepatitis B virus replicatie wordt beschreven.

In het tweede hoofdstuk beschrijven we de structuur en dynamica van de apical stem van het ε signaal voor inkapseling van het hepatitis B virus van de eend. Dit omdat de meeste moleculair biologische details over het replicatie mechanisme van het hepatitis B virus van de virus lijn van de eend komen. De in-vitro ε van de eend vertoont zowel bindings competentie met het protein P complex en DNA primer synthese, terwijl de humane ε alleen binding vertoont met het protein P complex. Deze structuur werd bepaald door een grote set ‘klassieke’ NMR restraints (NOE-afstands restraints en RDC restraints) mee te nemen in een structuurberekening. We hebben informatie over lokale dynamica afgeleid uit spinrelaxatie en relaxatiedispersie metingen. Hieruit hebben we gevonden dat de structuur ‘enhanced’ basenpaar opening vertoont, wat cruciaal is voor binding van proteinen en dus voor de functie van dit domein van het RNA molecuul.

In het derde en vierde hoofdstuk worden NMR studies aan het ε signaal voor inkapseling van het hepatitis B virus beschreven. Een beter begrip van de ε RNA structuur en daarmee van zijn functie, zijn essentieel om medicijnen te ontwerpen tegen chronische hepatitis B infectie. Eerst wordt de toekennings strategie voor de NMR spectra van de primer bulge van het ε signaal voor inkapseling van het hepatitis B virus van de eend gegeven. Daarna, werd de structuur en dynamica van de volledige ε van zowel de mens als de eend bepaald met behulp van NOE-afstands informatie, RDCs, en spin-relaxatie data. RDCs werden gemeten aan een sample met pf1 fagen als oriënterend medium. De RDCs werden gebruikt om domein bewegingen van de helices van de volledige ε te visualiseren. Zowel de ε structuur van de eend als van de mens maken een scharnier-beweging met een openingshoek van ~20°. Uit verdere dynamica studies volgt dat de primer synthese site vrij ligt van tijd tot tijd, en dat suggereert dat de weg dan vrij is voor primer synthese in de active site van het P-protein, die verantwoordelijk is voor genoom replicatie.

In het vijfde en zesde hoofdstuk, wordt de relatie tussen base positionering in helices nucleine zuren en proton chemische verschuiving vastgesteld. Dit heeft geleid tot een nieuw rekenprotocol die nucleine zuur helices uitrekent op basis van isotrope niet-uitwisselbare proton chemische verschuivingen. Er wordt gebruik gemaakt van het feit dat kringstroom effecten in de aromatische basen, dominant zijn in de chemische verschuiving van protonen van nucleine zuren. Het protocol, genaamd CHEOPS, optimaliseert de isotrope chemische verschuiving van protonen door het aanpassen van de oriëntatie en positionering van de basenvlakken in een helix door middel van een iteratieve singular value decompositie. Deze singular value decompositie maakt het protocol geschikt voor structuurbepaling met een incomplete lijst aan proton chemische verschuivingswaarden. De positie van de basevlakken kan bepaald worden tot 0.6 Å nauwkeurig, en de oriëntatie tot 6° nauwkeurig. Na positionering en orientatie van de basenvlakken, wordt het molecuul aan een moleculaire dynamica simulatie onderworpen om kleine afwijkingen in de backbone te corrigeren. Op deze wijze kunnen structuren precies en acuraat bepaald worden door een lijst met minimaal vier chemische verschuivingen per basenpaar. Naast een diepgaande analyse van de validiteit en toepasbaarheid van CHEOPS, wordt de methode ook gedemonstreerd op twee reeds bekende RNA structuren. De structuren werden opnieuw berekend aan de hand van alleen hun proton chemische verschuivingen.

In het zesde hoofdstuk, werd de methode gebruikt om de Dickseron dodecameer B-DNA helix structuur *de novo* te bepalen met chemische verschuivingen alleen. Onze methode was in staat om de structuur te reproduceren die voldoet aan zowel alle NMR restraints (zoals NOE en RDCs) als isotrope proton chemische verschuivingen. De NMR afgeleide structuur heeft veel gelijkenissen met de röntgen-diffractie structuren. Dit onderstreept de kwaliteit van onze methode, omdat bekend is dat de kwaliteit van röntgen-diffractie structuren beter is dan de kwaliteit van ‘klassieke’ NMR structuren.

Het onderzoek in de hoofdstukken twee tot vier heeft geresulteerd in de structuur en dynamica van een biologisch relevant viraal RNA, die mogelijk een sleutel biedt voor medicijnontwerp tegen chronische hepatitis B virus infectie. In het vierde hoofdstuk laten we een methode zien om faag-geinduceerde RDCS te gebruiken om de moleculaire dynamica van nucleine zuur domeinen te visualizeren. Dit, gecomplementeerd met hoofdstukken vijf en zes, waar we een methode beschrijven om nucleine helix structuren te berekenen op grond van hun proton chemische verschuiving, introduceert een varieteit aan nieuwe methoden om structuur en dynamica van RNA op te lossen met behulp van NMR.

Vooruitzichten

In dit proefschrift werden de structuur en dynamica opheldering van het ε element van het hepatitis B virus beschreven. Verder hebben we een hypothese geformuleerd over de binding van het ε element aan het reverse transcriptase domein van het P-protein. Verdere studies aan het complex van protein-P met ε zouden kunnen worden gebruikt om dit model en zijn details nader te onderzoeken. Voor deze studies is het nodig om zowel het vrije protein-P te bestuderen, als het complex tussen protein-P en ε RNA.

Als we de methodologische aspecten in dit proefschrift beschouwen, hebben we een nieuwe methode beschreven om nucleïne zuur helix structuren te bepalen op grond van alleen hun proton chemische verschuivingen. Verder onderzoek is nodig voordat deze methode kan worden gebruikt voor het bepalen van niet-helix elementen, zoals bulges en loops. Verder rijst de vraag hoe deze methode gebruik kan maken van heteronucleaire chemische verschuivingen, zoals ^{13}C, ^{15}N, en ^{31}P. Daarnaast zijn er nog studies nodig om structuren te bepalen met een combinatie van isotrope chemische verschuivingen en andere restraints, zoals RDCs en NOEs.

Het meenemen van dynamica is in deze methode ook een onderwerp voor verdere studies. Als de theorie hiervoor door zou worden ontwikkeld, zou men zowel dynamica en structuur informatie kunnen halen uit chemische verschuivingen alleen.

Het voorgestelde protocol in hoofdstukken vijf en zes opent de weg naar een nieuwe manier van structuurbepaling, zodra deze methode verder is ontwikkeld, kan deze methode een manier zijn om nog grotere NMR structuren acuraat en precies te bepalen dan nu mogelijk is met NMR spectroscopie. De methode zou ook gebruikt kunnen worden om structuren van toestanden van nucleïne zuren met korte levensduur en een lage populatie te bepalen, en zouden daarmee een beter inzicht kunnen geven in de rol van deze kort levende, laag bezette toestanden in de functie van het molecuul.

Chapter 8

Acknowledgements

Dankwoord

8 Acknowledgements / Dankwoord

Dit werk zou niet mogelijk zijn geweest zonder een hele serie mensen die aan mijn zijde stonden gedurende de ‘evolutie’ van dit proefschrift.

Sybren Wijmenga, bedankt voor je goede begeleiding. Je bent iemand die tijd neemt voor zijn promovendi, en iemand met altijd een optimistische noot. Marco Tessari, bedankt voor alle praktische hulp en het stellen van goede vragen in onze discussies.

Frank Nelissen, naast dat je gewoon een toffe collega bent, ben je onmisbaar geweest in mijn onderzoek. Bedankt voor het bereiden van een aantal samples, en alle testjes die je tussendoor hebt gedaan. Frederic Girard, bedankt voor je hulp en je geduld in de opstartfase van mijn onderzoek. Veel succes met Spinnovation, en natuurlijk met veel geluk met je carrierestap tot vader! Kirsten Ampt, ik ben blij dat je binnen het biomoleculaire veld ons kwam versterken als ervaren wetenschapper. Ik vond je aanwezigheid altijd erg waardevol, en ook leuk natuurlijk. Hugo van Ingen, al was het genoegen kort, ik heb toch het een en ander van je opgepikt. Veel succes met je verdere carriere. Otmar Ottink, bedankt dat ik je zo nu en dan van je werk mocht houden met vraagjes over details aangaande de wondere wereld van de NMR. Geerten Vuister, bedankt voor je commentaren op mijn presentaties. Ik heb deze als erg waardevol beschouwd. Cees Hilbers, bedankt voor je interesse in mijn onderzoek. Ik waardeer altijd heel erg dat je op een vriendelijke wijze hele kritische vragen weet te stellen. Daarnaast vond ik het fijn dat je tijd voor me nam voor kleinigheidjes, ondanks dat je gepensioneerd bent. Ik heb je aanwezigheid zeker gewaardeerd, en je positieve woorden waren een steun toen de natuur me tegen werkte. Hans Heus, ik ben erg blij met de enorme literatuurkennis die je hebt. Als ik een wonderlijke RNA structuur wilde vinden, was een ‘inval’ op jouw kantoortje vaak de moeite waard. Verder natuurlijk bedankt voor het nalezen van de manuscripten over chemical shift based structure derivation. Arno Kentgens, bedankt voor je onderwijs ik heb er veel aan gehad, en daarnaast is het leuk geweest om op die wijze een kijkje te nemen in de keuken van de vaste stof NMR. Michael Petersen, thank you for a good collaboration and thanks for your help on difficulties in Xplor. Janus Zdunek, thanks for your help on the re-compilation of a modified version of Xplor. Margriet Verkuijlen, bedankt dat ik jou zo nu en dan kon lastig vallen met kleinigheidjes over de wiskunde of de fysica. De skate-events waren altijd erg leuk. Paul Knijn, Je bent echt een hele toffe collega, en ik heb de movie-nights als een waardevolle aanvulling op het werk beschouwd. Frank ter Elst, je bent bij ons geweest als student en hebt mogen ervaren dat het in de wetenschap soms flink kan tegenzitten. Ik hoop desondanks dat je ergens een geschikte plek hebt gevonden, en dat je spoedig een bul in ontvangst mag nemen. Anna Andreychenko, thank you for your feedback on my work, and thanks for allowing me to bother you with small questions on physics. Ivo Westerweele, je hebt goed werk geleverd met de toekenning van de terminator stem van the guanine-sensing riboswitch. Ard Kolkman, eindelijk verlost van het geschop onder de tafel! Desondanks, leuk om je als collega te hebben gehad. Agnieszka Smolinska, you are a nice and kind colleague to work with. Beside that you are able to explain well what you are doing, thanks for all your help with different aspects of singular value decomposition. Marc Jupin, our Mauritian, I always enjoyed seeing your direct approach to problems. I especially liked you kicking some students away from a table in the canteen because they were playing cards. And for the record: I did not kill the dodo, although I do still wonder what it would taste like. Marian de With en Maaike Noortman, bedankt voor al jullie praktische hulp en werkzaamheden die me veel tijd bespaard hebben. Jos Joordens, de legende, bedankt voor je onderhoud aan computers, en bedankt voor het opvangen van de stommiteiten die je als computer gebruiker soms begaat. Jack van Asten, het genoegen was welliswaar kort maar ik heb je een prettige collega gevonden, hopelijk tot ziens in de MRI community. Jan van Os, Hans Janssen, Gerrit Janssen, bedankt voor het onderhoud van onze meet-apparatuur. Geen NMR apparaten, geen proefschrift he!? Chandrakala Gowda, I always enjoyed your company. You are always someone who gives others a smile. I also enjoyed the one time you cooked for us. Vipin Agarwal, it always makes me smile how you and Chandra disagree on many topics. I really enjoyed your company. Suresh Vasa, the big guy, it is always great to see how much you enjoy games and other things. I appreciate to be working with you. Andreas Brinkmann, Je bent altijd een bron van kennis geweest. Voor mij op het gebied van het Linux gebruik, en het gebruik van de cluster computer. Verder bewonder ik je snelle begrip en algemene interesse in zaken buiten jouw eigen werkveld. Jorge Villanueva-Garibay, I really enjoyed your company and was always amazed by your views on dutch politics. Aafke van der Vught, heerser over de ‘koffiepot’, en zonder koffie geen wetenschap, helemaal niet als er zoveel uren achter een computer worden doorgebracht. Gijs van der Heijden, het genoegen was kort, maar ik vond je een toffe collega. Dennis Grimminck, the one guy who really knows to tune

KDE! Veel succes met je onderzoek! Vincent Breukels, het is wel een gemis nu je in de NCMLS toren zitten. Jan Aelen.... Ik mis toch de felle discussies over de ziektekostenverzekeringen bij de koffiepauze, maar ik hoop dat je desondanks geniet van je pensioen. Bedankt voor uren van vermaak! Jisk Attema, fijn dat je onder het genot van een kop koffie zo nu en dan mij wilde aanhoren met het geklaag dat de natuur mij tegenwerkt.

Dan rest mij nog Ernst van Eck, Jan van Bentum, Eugenio Daviso, Anne-Jo Oosthoek-de Vries, Jacob Bart. En onze studenten: Nan Eshuis, Luuk Jansen, Rik Nievergeld, Jarno Verhoofstad, Lotte Wiermans, Anneke Mak, Bart Verberkt, Rik Sprenkels, Vivike Lapoutre, Jurjen Draaisma, Bas van Meerten, Leon van Sandwijk, Koen Tijssen, Ole Brauckmann, Robert-Jan Alers, Erik Stronks, en Liz Leunissen. Jullie allen bedankt voor jullie bijdrage aan een goede werksfeer.

Dan zijn er nog een aantal mensen buiten het veld van de wetenschap die zeker genoemd moeten worden. Mijn ouders, Bertus en Lida, omdat jullie me de mogelijkheid hebben geboden om mijn nieuwsgierigheid achterna te gaan, kan ik mede dankzij jullie de wereld verrijken met dit proefschrift. ‘Ome Rink’ en ‘Tante Liesbeth’, jullie bijstand in praktische regelzaken heeft mij bergen energie bespaard. Jantina en Herman Hartsuiker, bedankt voor het tijdelijk bieden van onderdak, en het spoedig ‘regelen’ van een andere woonruimte voor mij. Dit is een onmisbaar iets geweest in de afgelopen periode.

Frans van der Wielen en Luuk van der Wielen voor de support in de zogenaamde promotie-dip. Dan is er nog stam: Harmen de Jong, Hugo Buddelmeijer, Elsbeth Leeffers, Wieke Veer, Chinwei Ng, Jolanda Heida, Arjan Kuipers, Janine van Manen, Janneke Eppenga, Dirma ter Heege, Olof van der Berg, Jisk Attema (again) en Elena Crespo, iedereen hartelijk bedankt voor de tijdige, leuke en gezellige afleiding van promotieperikelen. En last maar zeer zeker not least, Marlous Bank, bedankt dat je de ontberingen hebt doorstaan met iemand als ik. Het zal niet altijd makkelijk zijn geweest, maar de verloren tijd gaan we snel inhalen. Bedankt dat je in deze tijd aan mijn zijde hebt gestaan, en bent blijven staan. Verder wil ik iedereen bedanken die op wat voor wijze dan ook een bijdrage aan dit werk geleverd hebben en die ik vergeten ben, al hoop ik dat dat aantal zich tot een minimum beperkt.

Ramon M. van der Werf

Chapter 9

Curriculum Vitae

9 Curriculum Vitae

Work experience

2011 – Date of Print	**MRI Physicist**, Erasmus Medical Centre, Rotterdam, The Netherlands Support for MRI users, and research for imaging of stem cell therapies
2009 – 2010	**Postdoc**, Leiden University, Leiden, The Netherlands Research concerning presence and stability of pseudotriloop structures in viral RNAs
2005 - 2009	**Junior researcher** (Ph.D. student), Radboud University Nijmegen, Nijmegen, The Netherlands Research for new methods to determine RNA structures by NMR spectroscopy
	Chemical analyst, laboratory of Friesland Dairy Foods, Leeuwarden, NL
Past-1998	A variety of other holiday jobs.

Education

2005-2009	WO	**Ph.D. Research** in Biophysical chemistry (NMR) Radboud University Nijmegen, Nijmegen, The Netherlands
1997-2004	WO	**Master in chemistry**, *Major: Biophysical chemistry* Minor: organic chemistry University of Groningen, Groningen, The Netherlands
1999-2000	WO	Psychology, University of Groningen, Groningen, The Netherlands
1995-1997	VWO	Bogerman, Sneek, The Netherlands
1994-1995	HAVO	Bogerman, Sneek, The Netherlands
1989-1994	MAVO	Bogerman College, Sneek, The Netherlands

Courses

2011	Advanced MRS Course, General Electric, Tenerife, Spain
2011	Pulse sequence programming in EPIC, General Electric, Italy
2011	Translational Imaging Workshop, Erasmus MC, The Netherlands
2010	Certificate Effective leadership, Red Cross, The Netherlands
2007	EMBO Workshop: Viral RNA; Structure & Function
2007	EU NMR Training workshop, Tarragona, Spain
2006-2007	Course communicating in German (Language level: A2)
2006	Course in academic writing
2006	BNRA: Advanced NMR; from data to structure
2005	BNRA: Advanced Nuclear Magnetic Resonance Spectroscopy
2003	*Certificate* in working with genetically modified organisms
1998	*Certificate* in presentation techniques

Publications

R.M. van der Werf, F.C. Girard, F. Nelissen, M. Tessari, S.S. Wijmenga (2008): *^{1}H, ^{13}C and ^{15}N NMR assignments of Duck HBV primer loop of the encapsidation signal epsilon*, Biomol. NMR Assignments, **2**(2), pp. 143-145

R.M. van der Werf, M. Tessari, S.S. Wijmenga (2011): Nucleic Acid Helix Structure Determination from NMR Proton Chemical Shifts, Under Revision

K.A.M. Ampt, R.M. van der Werf, F.H.T. Nelissen, M. Tessari, S.S. Wijmenga (2009): *The Unstable Part of the Apical Stem of Duck Hepatitis B Virus Epsilon Shows Enhanced Base Pair Opening but Not Pico- to Nanosecond Dynamics and Is Essential for Reverse Transcriptase Binding*, Biochemistry, **48**(44), pp.10499-10508

R.M. van der Werf, M. Tessari, S.S. Wijmenga (2011): *Dickerson Dodecamer Structure derived from NMR Proton Chemical Shifts*, Manuscript in preparation

R.M. van der Werf, M.I.P. Petersen, K.A.M. Ampt, J. Schleucher, S. Flodell, F. Girard, F. Nelissen, M. Tessari, M. Nassal, and S.S. Wijmenga (2010): *The structural basis for the initiation of the reverse transcription of the pregenomic RNA of the human and avian Hepatitis B virus. Comparison of the epsilon RNA elements*, Manuscript in preparation

Grants: EUROMAR grant, Tarragona, Spanje, juli 2007